CRFE

Access DB# 88325

SEARCH REQUEST FORM

Scientific and Technical Information Center

LB

Requester's Full Name: M.A. WALICKA Examiner #: 78201 Date: March 6, 03

Art Unit: 1652 Phone Number 305-7270 Serial Number: 09/554,414

Mail Box and Bldg/Room Location: 10D06 (10D01) Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need. mej

****

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: DNA demethylase, therapeutic & diagnostic etc (P2 P7)

Inventors (please provide full names): MOSHE SZYF et. al.

Earliest Priority Filing Date: May 11 / 1998

***For Sequence Searches Only* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.**

Please search

1. SEQ ID NO: 2 (02 pg 01)
   & ~~AA~~ 150 – 411 of SEQ ID NO: 2

2. The ~~following~~ nucleotide seq: (DNA)
   CGCGCGCG
   GCGCGCGC (06) walicka 414

Thank you in advance.

MWalicka

AA
2-411

RECEIVED
MAR - 6 2003
BIOCHEM (STIC)

****

STAFF USE ONLY

Searcher: R. Schreiber
Searcher Phone #: 308-4292
Searcher Location: CM1 6A03
Date Searcher Picked Up:
Date Completed: 3/17
Searcher Prep & Review Time: 14
Clerical Prep Time:
Online Time: 17

Type of Search

NA Sequence (#) 1
AA Sequence (#) 2
Structure (#)
Bibliographic
Litigation
Fulltext
Patent Family
Other

Vendors and cost where applicable

STN
Dialog
Questel/Orbit
Dr.Link
Lexis/Nexis
Sequence Systems Compugen IG
WWW/Internet
Other (specify)

PTO-1590 (8-01)

GenCore version 5.1.4_p5_4578

OM protein - protein search, using sw model

Run on: March 12, 2003, 05:40:46 ; Search time 12.4577 Seconds
(without alignments)
618.801 Million cell updates/sec

Title: US-09-554-414B-2_COPY_150_411
Perfect score: 1344
Sequence: 1 MDCPALPPGWKKEEVIRKSG.....LSRAADTEEMDIEMDSGDEA 262

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1344	100.0	263	4	US-09-149-476-580	Sequence 580, App
2	214.5	16.0	574	4	US-09-079-431B-6	Sequence 6, Appli
3	214.5	16.0	629	4	US-09-079-431B-4	Sequence 4, Appli
4	214.5	16.0	630	4	US-09-079-431B-2	Sequence 2, Appli
5	126	9.4	580	4	US-09-327-984A-38	Sequence 38, Appl
6	97	7.2	448	3	US-08-476-509B-2	Sequence 2, Appli
7	96.5	7.2	694	3	US-08-559-397A-31	Sequence 31, Appl
8	95.5	7.1	1312	4	US-09-041-886-19	Sequence 19, Appl
9	93	6.9	1048	3	US-09-356-952-5	Sequence 5, Appli
10	92.5	6.9	459	2	US-08-870-518-1	Sequence 1, Appli
11	90.5	6.7	449	1	US-08-476-008-5	Sequence 5, Appli
12	90.5	6.7	449	1	US-08-476-008-7	Sequence 7, Appli
13	90.5	6.7	449	1	US-08-306-063-5	Sequence 5, Appli
14	90.5	6.7	449	1	US-08-306-063-7	Sequence 7, Appli
15	90.5	6.7	449	1	US-08-833-485-5	Sequence 5, Appli
16	90.5	6.7	449	1	US-08-833-485-7	Sequence 7, Appli
17	90.5	6.7	449	4	US-09-137-440-5	Sequence 5, Appli
18	90.5	6.7	449	4	US-09-137-440-7	Sequence 7, Appli
19	90.5	6.7	449	5	PCT-US91-06148A-5	Sequence 5, Appli
20	90.5	6.7	449	5	PCT-US91-06148A-7	Sequence 7, Appli
21	88.5	6.6	1070	4	US-09-091-042A-2	Sequence 2, Appli
22	88	6.5	465	5	PCT-US93-08386-5	Sequence 5, Appli
23	88	6.5	486	3	US-08-348-518C-2	Sequence 2, Appli
24	88	6.5	507	5	PCT-US93-08386-8	Sequence 8, Appli
25	88	6.5	10182	4	US-09-134-001C-3159	Sequence 3159, Ap
26	87	6.5	861	1	US-08-484-105-18	Sequence 18, Appl
27	87	6.5	861	1	US-08-484-106-18	Sequence 18, Appl
28	87	6.5	2285	4	US-09-308-375-2	Sequence 2, Appli
29	86.5	6.4	459	2	US-08-870-518-2	Sequence 2, Appli
30	86.5	6.4	544	3	US-08-559-397A-30	Sequence 30, Appl
31	85.5	6.4	612	1	US-08-344-695-2	Sequence 2, Appli
32	85.5	6.4	1404	4	US-08-801-308-1	Sequence 1, Appli
33	85.5	6.4	1792	2	US-08-962-284-4	Sequence 4, Appli
34	84	6.2	454	3	US-08-348-518C-4	Sequence 4, Appli
35	84	6.2	454	3	US-08-476-509B-4	Sequence 4, Appli
36	82.5	6.1	578	4	US-09-066-046-6	Sequence 6, Appli
37	82.5	6.1	578	4	US-08-975-762-50	Sequence 50, Appl
38	82.5	6.1	578	4	US-09-295-028-50	Sequence 50, Appl
39	82.5	6.1	578	4	US-09-106-582-50	Sequence 50, Appl
40	82.5	6.1	919	2	US-08-588-983-9	Sequence 9, Appli
41	82.5	6.1	919	2	US-08-588-976-9	Sequence 9, Appli
42	81.5	6.1	427	3	US-09-196-857-2	Sequence 2, Appli
43	81.5	6.1	801	1	US-08-725-012-2	Sequence 2, Appli
44	81.5	6.1	3696	4	US-09-134-001C-5080	Sequence 5080, Ap
45	80.5	6.0	680	3	US-08-947-965-77	Sequence 77, Appl

ALIGNMENTS

RESULT 1
US-09-149-476-580
; Sequence 580, Application US/09149476
; Patent No. 6420526
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 186 Human Secreted proteins
; FILE REFERENCE: PZ002P1
; CURRENT APPLICATION NUMBER: US/09/149,476
; CURRENT FILING DATE: 1998-09-08
; EARLIER APPLICATION NUMBER: PCT/US98/04493
; EARLIER FILING DATE: 1998-03-06
; EARLIER APPLICATION NUMBER: 60/040,162
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,333
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/038,621
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,626
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,334
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,336
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,163
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/047,600
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,615
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,597
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,502
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,633
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,583
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,617
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,618
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,503
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,592
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,581
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,584
; EARLIER FILING DATE: 1997-05-23

; EARLIER APPLICATION NUMBER: 60/047,500
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,587
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,492
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,598
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,613
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,582
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,596
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,612
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,632
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,601
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/043,580
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,568
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,314
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,569
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,311
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,671
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,674
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,669
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,312
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,313
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,672
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,315
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/048,974
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/056,886
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,877
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,889
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,893
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,630
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,878
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,662
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,872
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,882
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,637
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,903
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,888
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,879
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,880
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,894
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,911
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,636
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,874
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,910
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,864
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,631
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,845
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,892
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/057,761
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/047,595
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,599
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,588
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,585
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,586
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,590
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,594
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,589
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,593
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,614
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/043,578
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,576
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/047,501
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/043,670
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/056,632
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,664
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,876
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,881
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,909
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,875
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,862
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,887
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,908
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/048,964
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/057,650
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: 60/056,884
; EARLIER FILING DATE: 1997-08-22

```
;  EARLIER APPLICATION NUMBER: 60/057,669
;  EARLIER FILING DATE: 1997-09-05
;  EARLIER APPLICATION NUMBER: 60/049,610
;  EARLIER FILING DATE: 1997-06-13
;  EARLIER APPLICATION NUMBER: 60/061,060
;  EARLIER FILING DATE: 1997-10-02

  Query Match             100.0%;  Score 1344;  DB 4;  Length 263;
  Best Local Similarity   100.0%;  Pred. No. 7.6e-123;
  Matches  262;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 MDCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNTVDLSSFDFR 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MDCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNTVDLSSFDFR 60

Qy     61 TGKMMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKVKS 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TGKMMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKVKS 120

Qy    121 DPQRMNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSAVASAL 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 DPQRMNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSAVASAL 180

Qy    181 HTSSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEALMA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 HTSSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEALMA 240

Qy    241 DILSRAADTEEMDIEMDSGDEA 262
          ||||||||||||||||||||||
Db    241 DILSRAADTEEMDIEMDSGDEA 262


RESULT 2
US-09-079-431B-6
; Sequence 6, Application US/09079431B
; Patent No. 6326484
; GENERAL INFORMATION:
;  APPLICANT: Gage, Fred
;  APPLICANT:  Ueba, Tetsuya
;  TITLE OF INVENTION: TRANSCRIPTION FACTOR REGULATING FGF-2
;  TITLE OF INVENTION:  AND VARIANTS THEREOF
;  FILE REFERENCE: SALKINS.015A
;  CURRENT APPLICATION NUMBER: US/09/079,431B
;  CURRENT FILING DATE:  1998-05-14
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6
;   LENGTH: 574
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-079-431B-6

  Query Match             16.0%;  Score 214.5;  DB 4;  Length 574;
  Best Local Similarity   45.7%;  Pred. No. 1.7e-12;
  Matches   42;  Conservative   13;  Mismatches   30;  Indels    7;  Gaps    1;

Qy      1 MDCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNTVDLSSFDFR 60
          :||||| ||||: || |||| : |:|| || ||:| : ||| :| ||||   ||: |||:
Db      6 LDCPALGPGWKRREVFRKSGATCGRSDTYYQSPTGDRIRSKVELTRYLGPACDLTLFDFK 65

Qy     61 TG-------KMMPSKLQKNKQRLRNDPLNQNK 85
          |       |  |  :   |::  : |    |
Db     66 QGILCYPAPKAHPVAVASKKRKKPSRPAKTRK 97


RESULT 3
US-09-079-431B-4
; Sequence 4, Application US/09079431B
; Patent No. 6326484
; GENERAL INFORMATION:
;  APPLICANT: Gage, Fred
;  APPLICANT:  Ueba, Tetsuya
;  TITLE OF INVENTION: TRANSCRIPTION FACTOR REGULATING FGF-2
;  TITLE OF INVENTION:  AND VARIANTS THEREOF
;  FILE REFERENCE: SALKINS.015A
;  CURRENT APPLICATION NUMBER: US/09/079,431B
;  CURRENT FILING DATE:  1998-05-14
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
;   LENGTH: 629
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-079-431B-4

  Query Match             16.0%;  Score 214.5;  DB 4;  Length 629;
  Best Local Similarity   45.7%;  Pred. No. 1.9e-12;
  Matches   42;  Conservative   13;  Mismatches   30;  Indels    7;  Gaps    1;

Qy      1 MDCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNTVDLSSFDFR 60
          :||||| ||||: || |||| : |:|| || ||:| : ||| :| ||||   ||: |||:
Db      6 LDCPALGPGWKRREVFRKSGATCGRSDTYYQSPTGDRIRSKVELTRYLGPACDLTLFDFK 65

Qy     61 TG-------KMMPSKLQKNKQRLRNDPLNQNK 85
          |       |  |  :   |::  : |    |
Db     66 QGILCYPAPKAHPVAVASKKRKKPSRPAKTRK 97


RESULT 4
US-09-079-431B-2
; Sequence 2, Application US/09079431B
; Patent No. 6326484
; GENERAL INFORMATION:
;  APPLICANT: Gage, Fred
;  APPLICANT:  Ueba, Tetsuya
;  TITLE OF INVENTION: TRANSCRIPTION FACTOR REGULATING FGF-2
;  TITLE OF INVENTION:  AND VARIANTS THEREOF
;  FILE REFERENCE: SALKINS.015A
;  CURRENT APPLICATION NUMBER: US/09/079,431B
;  CURRENT FILING DATE:  1998-05-14
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 630
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-079-431B-2

  Query Match             16.0%;  Score 214.5;  DB 4;  Length 630;
  Best Local Similarity   45.7%;  Pred. No. 1.9e-12;
  Matches   42;  Conservative   13;  Mismatches   30;  Indels    7;  Gaps    1;

Qy      1 MDCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNTVDLSSFDFR 60
          :||||| ||||: || |||| : |:|| || ||:| : ||| :| ||||   ||: |||:
Db      6 LDCPALGPGWKRREVFRKSGATCGRSDTYYQSPTGDRIRSKVELTRYLGPACDLTLFDFK 65

Qy     61 TG-------KMMPSKLQKNKQRLRNDPLNQNK 85
          |       |  |  :   |::  : |    |
Db     66 QGILCYPAPKAHPVAVASKKRKKPSRPAKTRK 97


RESULT 5
US-09-327-984A-38
; Sequence 38, Application US/09327984A
; Patent No. 6368594
; GENERAL INFORMATION:
;  APPLICANT:  Doetsch, Paul W.
;  APPLICANT:  Kaur, Balveen
;  APPLICANT:  Avery, Angela M.
;  TITLE OF INVENTION:  Broad Specificity DNA Damage Endonuclease
;  FILE REFERENCE:  25-98
;  CURRENT APPLICATION NUMBER: US/09/327,984A
;  CURRENT FILING DATE:  1999-06-08
;  PRIOR APPLICATION NUMBER:  US 60/088,521
;  PRIOR FILING DATE:  1998-06-08
;  PRIOR APPLICATION NUMBER:  US 60/134,752
```

```
;  PRIOR FILING DATE:  1999-05-18
;  NUMBER OF SEQ ID NOS:  39
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 38
;   LENGTH: 580
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-327-984A-38

  Query Match             9.4%;  Score 126;  DB 4;  Length 580;
  Best Local Similarity   22.3%;  Pred. No. 0.00068;
  Matches   72;  Conservative   44;  Mismatches  109;  Indels   98;  Gaps   13;

Qy       5 ALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNTVDLS----SFDFR 60
           ::| ||::    |  | :||: |||: || | |||||  || ||    : |     |||
Db      85 SVPCGWERVVKQRLFGKTAGRFDVYFISPQGLKFRSKSSLANYLHKNGETSLKPEDFDFT 144

Qy      61 T-------GKMMPSKLQKNKQRLRNDPLNQN-------KGKPDL---------------- 90
                  :     :       |:|   | |       | | |:
Db     145 VLSKRGIKSRYKDCSMAALTSHLQNQSNNSNWNLRTRSKCKKDVFMPPSSSSELQESRGL 204

Qy      91 ----NTTLPIRQTASI-------FKQPVTKVT---NHPSNKVKSDPQR------MNEQPR 130
               :| | :::   :         ::|  |||     |  | |   ::       ::  |
Db     205 SNFTSTHLLLKEDEGVDDVNFRKVRKPKGKVTILKGIPIKKTKKGCRKSCSGFVQSDSKR 264

Qy     131 QLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSAVASALHTSSAPITGQ 190
           :    |       |   :| : :  :| : :     |:  |||               :| :
Db     265 ESVCNK------ADAESEPVAQKSQLDRTVCISDAGACGETL------------SVTSE 305

Qy     191 VSAAVEK---------NPAVWLNTSQPLCKAFIVTDEDIRKQEE----------RVQQV 230
           :: |:|         |      ||  : |     | :  :: |           :|: |
Db     306 ENSLVKKKERSLSSGSNFCSEQKTSGIINKFCSAKDSEHNEKYEDTFLESEEIGTKVEVV 365

Qy     231 RKKLEEALMADILSRAADTEEMD 253
           :|  | |  ||| | :   |||
Db     366 ERK--EHLHTDILKRGS---EMD 383


RESULT 6
US-08-476-509B-2
; Sequence 2, Application US/08476509B
; Patent No. 6034212
;  GENERAL INFORMATION:
;    APPLICANT:  SUDOL, MARIUS
;    APPLICANT:  PEER, BORK
;    APPLICANT:  HENRY, CHEN
;    TITLE OF INVENTION:  A SH2 DOMAIN ASSOCIATED PROTEIN, A
;    TITLE OF INVENTION:  SIGNALLING DOMAIN THEREIN, NUCLEIC ACIDS ENCODING THE
;    TITLE OF INVENTION:  PROTEIN AND THE DOMAIN, AND DIAGNOSTIC AND THERAPEUTICUSES
;    TITLE OF INVENTION:  THEREOF
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Klauber & Jackson
;      STREET:  411 Hackensack Avenue
;      CITY:  Hackensack
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/476,509B
;      FILING DATE:  01-DEC-1994
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jackson Esq., David A.
;      REGISTRATION NUMBER:  26,742
;      REFERENCE/DOCKET NUMBER:  600-1-101 CIP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201 487-5800
;      TELEFAX:  201 343-1684
;      TELEX:  133521
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  448 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-476-509B-2

  Query Match             7.2%;  Score 97;  DB 3;  Length 448;
  Best Local Similarity   22.0%;  Pred. No. 0.31;
  Matches   64;  Conservative   35;  Mismatches  114;  Indels   78;  Gaps   10;

Qy       2 DCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNTVDLSSFDFRT 61
           | | |||||:  :                :|||:         ||  | :| ::
Db     168 DVP-LPPGWEMAK----------------TPSGQ---------RYFLNHIDQTTTWQDP 200

Qy      62 GKMMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKV--- 118
           | | |  | |      : |: ||           | |:| : |||       |   |
Db     201 RKAMLS--QMNVTAPTSPPVQQNLMNSASAMNQRISQSAPV-KQPPPLAPQSPQGGVMGG 257

Qy     119 KSDPQRMNEQPRQLFWEKRLQGLSASDVTEQ------IIKTMELPKGLQG--VGPG---- 166
           |  |:   : :||  ||    |   ::  |       : |||   | |     ||
Db     258 SSSNQQQQMRLQQLQMEKERLRLKHQELLRQELALRSQLPTMEQDGGSQNPVSSPGMSQE 317

Qy     167 -------SNDETLLSAVASALHTSSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDED 219
                 |:|  | |    :   |:       | :| : |  :||:   :     ::  :
Db     318 LRTMTTNSSDPFLNSGTYHSRDESTDSGLSMSSYSVPRTPDDFLNSVDEMDTGDSISQSN 377

Qy     220 IRKQEERV-------------------------QQVRKKLEEALMADILS 244
           |   : |                         :::   |:||| :|||:
Db     378 IPSHQNRFPDYLEAIPGTNVDLGTLEGDGMNIEGEELMPSLQEALSSDILN 428


RESULT 7
US-08-559-397A-31
; Sequence 31, Application US/08559397A
; Patent No. 6083713
;  GENERAL INFORMATION:
;    APPLICANT:  Manly, Susan P.
;    APPLICANT:  Kozlowski, Michael R.
;    APPLICANT:  Neve, Rachael L.
;    TITLE OF INVENTION:  CLONING AND EXPRESSION OF
;    TITLE OF INVENTION:  BETA APP-C100 RECEPTOR (C100-R)
;    NUMBER OF SEQUENCES:  35
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10036/2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/559,397A
;      FILING DATE:  15-NOV-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Coruzzi, Laura A
;      REGISTRATION NUMBER:  30,742
;      REFERENCE/DOCKET NUMBER:  6013-135
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-790-9090
;      TELEFAX:  212-869-8864
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  31:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  694 amino acids
```

```
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
US-08-559-397A-31

  Query Match             7.2%;  Score 96.5;  DB 3;  Length 694;
  Best Local Similarity   21.6%;  Pred. No. 0.65;
  Matches   65;  Conservative   36;  Mismatches   89;  Indels  111;  Gaps   12;

Qy       6 LPPGWKKEEVIRKSGLSAGK--------------SDVYYFSPSGKKFRS---------K 41
           ||  |:|||    ||:|  :              ||    :   | |::
Db     141 LPEEWEKEET--SSGISKREQQQNMQAVMDIVKFYQDVTETNGEDKMFKTFNTTTGLPGS 198

Qy      42 PQLA--------RYLGNTVDLSSFDFRTGKMMPSKLQKNKQRLRNDPLNQNKGKPDLNTT 93
           ||::        ::  :| |  ::  |||                 |:: :   | |||
Db     199 PQVSTPPANSFNKFPPSTSDSHNYGSRTG---------------TPMSNHVMSPTLNTD 242

Qy      94 --------LPIR----------QTASIFKQPV---------TKVTNHPSNKVKSDPQRMN 126
                  :| |           :| | | ||          | |: |:   :: | |
Db     243 SSSANGKFIPSRPAPKPPSSASASAPIIKSPVMNSAANVSPLKQTHAPTTPNRTSPNR-- 300

Qy     127 EQPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSAVASALHTSSAP 186
                               | :   | |  | : |    ::  | : |  ||
Db     301 --------------------SSISRNATLKKEEQPLPPIPPTKSKTSPIISTAHTP--- 336

Qy     187 ITGQVSAAVEKNPA---VWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEALMADIL 243
             |  |   | ||   |   ||:|     :  : : :|:|||  : |||   | : :|
Db     337 ---QQVAQSPKAPAQETVTTPTSKPAQARSLSKELNEKKREER--ERRKKQLYAKLNEIC 391

Qy     244 S 244
           |
Db     392 S 392


RESULT 8
US-09-041-886-19
; Sequence 19, Application US/09041886
; Patent No. 6235872
;  GENERAL INFORMATION:
;    APPLICANT:  Bredesen, Dale E.
;    APPLICANT:  Rabizadeh, Sharroz
;    TITLE OF INVENTION:  Proapoptotic Peptides, Dependence
;    TITLE OF INVENTION:  Polypeptides and Methods of Use
;    NUMBER OF SEQUENCES:  72
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell & Flores LLP
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/041,886
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-LJ 2626
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1312 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-041-886-19

  Query Match             7.1%;  Score 95.5;  DB 4;  Length 1312;
  Best Local Similarity   23.5%;  Pred. No. 2.1;
  Matches   43;  Conservative   38;  Mismatches   65;  Indels   37;  Gaps    9;

Qy      67 SKLQKNKQRLRNDPL-NQNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKVKSDPQRM 125
           |:||  :|   | |  |:   ||  | | |             :|  |   : | |: ::.
Db     748 SRLQDQRQ---NSPAGNKENIKP--NETSP----------SFSKAENKGISPVVSEHRKQ 792

Qy     126 NEQPRQLFWEKRLQGLSASDVTEQII-------KTMELPKGLQGVGPGSNDETLLSAVAS 178
            :  ::   : |||  | |:  :|::       |: :| |    : | : |  : :: ::
Db     793 IDDLKKFKNDFRLQPSSTSESMDQLLNKNREGEKSRDLIK--DKIEPSAKDSFIENSSSN 850

Qy     179 ALHTSSAPITGQVSAAV------EKNPAV---WLNTSQPLCKAFIVTDEDIRKQEERVQQ 229
             || | :  :| ::      :: | |    : || | ||      :|  :::  :|
Db     851 CTSGSSKPNSPSISPSILSNTEHKRGPEVTSQGVQTSSPACKQ---EKDDKEEKKDAAEQ 907

Qy     230 VRK 232
           |||
Db     908 VRK 910


RESULT 9
US-09-356-952-5
; Sequence 5, Application US/09356952
; Patent No. 6117663
; GENERAL INFORMATION:
;  APPLICANT: Boriack-Sjodin, Ann
;  APPLICANT: Margarit, S. M.
;  APPLICANT: Bor-Sogi, Dafna
;  APPLICANT: Cole, Philip
;  APPLICANT: Kuriyan, John
;  TITLE OF INVENTION: A CRYSTAL OF A RAS-SOS COMPLEX AND METHODS OF USE
;  TITLE OF INVENTION: THEREOF
;  FILE REFERENCE: 600-1-228N
;  CURRENT APPLICATION NUMBER: US/09/356,952
;  CURRENT FILING DATE: 1999-07-19
;  EARLIER APPLICATION NUMBER: 60/093,631
;  EARLIER FILING DATE: 1998-07-21
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
;   LENGTH: 1048
;   TYPE: PRT
;   ORGANISM: Saccharomyces cerevisiae
US-09-356-952-5

  Query Match             6.9%;  Score 93;  DB 3;  Length 1048;
  Best Local Similarity   21.7%;  Pred. No. 2.6;
  Matches   52;  Conservative   42;  Mismatches   90;  Indels   56;  Gaps   12;

Qy      54 LSSFDFRTGKMMPSKLQKNKQRL-------RNDPLNQN----------------KGKPD 89
           || |          | | : :: :        |  || |:                  |:
Db      85 LSRFSINANIWFHSTLFEEREAIASQKDPERRSPLLQSILGTFQKFHFLLRLLHFLSNPN 144

Qy      90 LNTTLPIRQTASIFKQPVTKVT-NHPSNKVKSDPQRMNEQPRQLFWEKRLQGLSASDVTE 148
             | || : |   ||      :: |:|  : : :     :: |||:   | : | |   | |
Db     145 ELTILP-QLTPRFFKDSFNTISWNNPFLRKRLNQHMSHDLPRQMI--KAVAGASGI-VAE 200

Qy     149 QIIKTMELPKGLQGVGPGSNDETLLSAVASALHTSSAP--------ITGQVSAAVEKNPA 200
             |    |:|   ||    |          :: |: |||         |   || : :::
Db     201 NI---DEIPASKQGTSCSSE---------TSHHSPSAPFQRRRRGTIFSNVSGSSDESDT 248

Qy     201 VWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEALMA--DILSRAADTEEM-DIEMD 257
           :|    :|    : : :| :     | :|:  ||:: : :  : ||  |:  :| : ||:
Db     249 IWSKRKKP----YPLNEETLSLVRARKKQLDGKLKQMIKSANEYLSNTANFSKMLNFEMN 304


RESULT 10
US-08-870-518-1
; Sequence 1, Application US/08870518
```

```
; Patent No. 5925566
;  GENERAL INFORMATION:
;    APPLICANT:  Davis, Roger J.
;    APPLICANT:  Galcheva-Gargova, Zoya
;    TITLE OF INVENTION:  NON-ACTIVATED RECEPTOR COMPLEX
;    TITLE OF INVENTION:  PROTEINS AND USES THEREOF
;    NUMBER OF SEQUENCES:  35
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  US
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows95
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/870,518
;      FILING DATE:  06-JUN-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/019,219
;      FILING DATE:  06-JUN-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fasse, Peter J.
;      REGISTRATION NUMBER:  32,983
;      REFERENCE/DOCKET NUMBER:  04020/102001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/542-5070
;      TELEFAX:  617/542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  459 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    FRAGMENT TYPE:  internal
US-08-870-518-1

  Query Match             6.9%;  Score 92.5;  DB 2;  Length 459;
  Best Local Similarity   22.9%;  Pred. No. 0.87;
  Matches   53;  Conservative   38;  Mismatches   81;  Indels   59;  Gaps   12;

Qy      33 PSGKKFRSKPQLARYLGNTVDLSSFDFRT---GKMMPSKLQKNKQR-----LRNDPLNQN 84
           |||  |   |  |    | : :: :| ||    :|:  : :  :::     |||: |  |
Db     199 PSGNSFVENPH-APQKDNALVITYYD-RTPQQAEMLGLQAEAPEEKAEEEDLRNEVLQFN 256

Qy      85 KGKPDLN----TTLPIRQTASIFKQPVTKVTN-----HPSNKVKS----DPQRMNEQPRQ 131
             |: |    | : : |    ||: :   ||     | :|:|||    :|
Db     257 TNCPECNAPAQTNMKLVQIPH-FKEVIIMATNCENCGHRTNEVKSGGAVEP--------- 306

Qy     132 LFWEKRLQGLSASDVTEQIIK----TMELPKGLQGVGPGSNDETLLSAVASALHTSSAPI 187
           |     |     ||:|  ::|    ::|:|:          :  |  ||     |:   :
Db     307 LGTRITLHITDPSDMTRDLLKSETCSVEIPE---------LEFELGMAVLGGKFTTLEGL 357

Qy     188 TGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEAL 238
             :   | |||             | : |     | |::|:  :|| : :
Db     358 LKDIRELVTKNP------------FTLGDSSNPDQSEKLQEFSQKLGQII 395


RESULT 11
US-08-476-008-5
; Sequence 5, Application US/08476008
; Patent No. 5627061
;  GENERAL INFORMATION:
;    APPLICANT:  Barry, Gerard F.
;    APPLICANT:  Kishore, Ganesh M.
;    APPLICANT:  Padgette, Stephen R.
;    APPLICANT:  Stallings, William C.
;    TITLE OF INVENTION:  Glyphosate Tolerant
;    TITLE OF INVENTION:  5-Enolpyruvylshikimate-3-Phosphate Synthases
;    NUMBER OF SEQUENCES:  69
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dennis R. Hoerner, Jr., Monsanto Co. BB4F
;      STREET:  700 Chesterfield Village Parkway
;      CITY:  St. Louis
;      STATE:  Missouri
;      COUNTRY:  USA
;      ZIP:  63198
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/476,008
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/306,063
;      FILING DATE:  13-SEP-1994
;      APPLICATION NUMBER:  US 07/749,611
;      FILING DATE:  28-AUG-1991
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/576,537
;      FILING DATE:  31-AUG-1990
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hoerner Jr., Dennis R.
;      REGISTRATION NUMBER:  30,914
;      REFERENCE/DOCKET NUMBER:  38-21(10660)A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (314)537-6099
;      TELEFAX:  (314)537-6047
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  449 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-476-008-5

  Query Match             6.7%;  Score 90.5;  DB 1;  Length 449;
  Best Local Similarity   22.0%;  Pred. No. 1.3;
  Matches   57;  Conservative   38;  Mismatches  123;  Indels   41;  Gaps    9;

Qy      16 IRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGN-------TVDL-SSFDFRTGKMMPS 67
           ||| |      ||:  :  |     :|: |   ||        |: |  ::| :|  :  :
Db      70 IRKEG------DVWIINGVGNGCLLQPEAALDFGNAGTGARLTMGLVGTYDMKTSFIGDA 123

Qy      68 KLQKNKQRLRNDPLNQNKGKPDL--NTTLPIRQTASIFKQPVTKVTNHPSNKVKSDPQRM 125
             | |       :|| :   : :      :|:          |:|      | :|||
Db     124 SLSKRPMGRVLNPLREMGVQVEAADGDRMPLTLIGPKTANPITYRVPMASAQVKS----- 178

Qy     126 NEQPRQLFWEKRLQGLSASDVT---EQIIKTMELPKGLQGVGPGSNDETLLSAVASALHT 182
                     | ||:   ||   | ::      | ||| |     ||    |     |
Db     179 ---------AVLLAGLNTPGVTTVIEPVMTRDHTEKMLQGFGADLTVETDKDGVRHIRIT 229

Qy     183 SSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEALMADI 242
             : || :  |  :|:   :|: ||  | :|   |:  :   :  |     |:  :
Db     230 GQGKLVGQ-TIDVPGDPS---STAFPLVAALLVEGSDVTIRNVLMNPTR----TGLILTL 281

Qy     243 LSRAADTEEMDIEMDSGDE 261
               || | ::  :  |::
Db     282 QEMGADIEVLNARLAGGED 300


RESULT 12
US-08-476-008-7
; Sequence 7, Application US/08476008
; Patent No. 5627061
;  GENERAL INFORMATION:
```

```
;    APPLICANT:  Barry, Gerard F.
;    APPLICANT:  Kishore, Ganesh M.
;    APPLICANT:  Padgette, Stephen R.
;    APPLICANT:  Stallings, William C.
;    TITLE OF INVENTION:  Glyphosate Tolerant
;    TITLE OF INVENTION:  5-Enolpyruvylshikimate-3-Phosphate Synthases
;    NUMBER OF SEQUENCES:  69
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dennis R. Hoerner, Jr., Monsanto Co. BB4F
;      STREET:  700 Chesterfield Village Parkway
;      CITY:  St. Louis
;      STATE:  Missouri
;      COUNTRY:  USA
;      ZIP:  63198
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/476,008
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/306,063
;      FILING DATE:  13-SEP-1994
;      APPLICATION NUMBER:  US 07/749,611
;      FILING DATE:  28-AUG-1991
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/576,537
;      FILING DATE:  31-AUG-1990
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hoerner Jr., Dennis R.
;      REGISTRATION NUMBER:  30,914
;      REFERENCE/DOCKET NUMBER:  38-21(10660)A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (314)537-6099
;      TELEFAX:  (314)537-6047
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  449 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-476-008-7

  Query Match             6.7%;  Score 90.5;  DB 1;  Length 449;
  Best Local Similarity   22.0%;  Pred. No. 1.3;
  Matches   57;  Conservative   38;  Mismatches  123;  Indels   41;  Gaps    9;

Qy      16 IRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGN-------TVDL-SSFDFRTGKMMPS 67
           ||| |      ||:  :  |     :|: |   ||       |: |  ::| :|  :  :
Db      70 IRKEG------DVWIINGVGNGCLLQPEAALDFGNAGTGARLTMGLVGTYDMKTSFIGDA 123

Qy      68 KLQKNKQRLRNDPLNQNKGKPDL--NTTLPIRQTASIFKQPVTKVTNHPSNKVKSDPQRM 125
             | |        :|| :   : :      :|:         |:|       | :|||
Db     124 SLSKRPMGRVLNPLREMGVQVEAADGDRMPLTLIGPKTANPITYRVPMASAQVKS----- 178

Qy     126 NEQPRQLFWEKRLQGLSASDVT---EQIIKTMELPKGLQGVGPGSNDETLLSAVASALHT 182
                       | ||:   ||   | ::      | ||| |     ||    |     |
Db     179 ---------AVLLAGLNTPGVTTVIEPVMTRDHTEKMLQGFGADLTVETDKDGVRHIRIT 229

Qy     183 SSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEALMADI 242
           : || :  |  :|:   :|: ||  | :|   |:  :   :   |      |:  :
Db     230 GQGKLVGQ-TIDVPGDPS---STAFPLVAALLVEGSDVTIRNVLMNPTR----TGLILTL 281

Qy     243 LSRAADTEEMDIEMDSGDE 261
             || | ::  :  |::
Db     282 QEMGADIEVLNARLAGGED 300
```

```
RESULT 13
US-08-306-063-5
; Sequence 5, Application US/08306063
; Patent No. 5633435
;  GENERAL INFORMATION:
;    APPLICANT:  Barry, Gerard F.
;    APPLICANT:  Kishore, Ganesh M.
;    APPLICANT:  Padgette, Stephen R.
;    APPLICANT:  Stallings, William C.
;    TITLE OF INVENTION:  Glyphosate Tolerant
;    TITLE OF INVENTION:  5-Enolpyruvylshikimate-3-Phosphate Synthases
;    NUMBER OF SEQUENCES:  69
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dennis R. Hoerner, Jr., Monsanto Co. BB4F
;      STREET:  700 Chesterfield Village Parkway
;      CITY:  St. Louis
;      STATE:  Missouri
;      COUNTRY:  USA
;      ZIP:  63198
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/306,063
;      FILING DATE:  13-SEP-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/749,611
;      FILING DATE:  28-AUG-1991
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/576,537
;      FILING DATE:  31-AUG-1990
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hoerner Jr., Dennis R.
;      REGISTRATION NUMBER:  30,914
;      REFERENCE/DOCKET NUMBER:  38-21(10660)A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (314)537-6099
;      TELEFAX:  (314)537-6047
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  449 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-306-063-5

  Query Match             6.7%;  Score 90.5;  DB 1;  Length 449;
  Best Local Similarity   22.0%;  Pred. No. 1.3;
  Matches   57;  Conservative   38;  Mismatches  123;  Indels   41;  Gaps    9;

Qy      16 IRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGN-------TVDL-SSFDFRTGKMMPS 67
           ||| |      ||:  :  |     :|: |   ||       |: |  ::| :|  :  :
Db      70 IRKEG------DVWIINGVGNGCLLQPEAALDFGNAGTGARLTMGLVGTYDMKTSFIGDA 123

Qy      68 KLQKNKQRLRNDPLNQNKGKPDL--NTTLPIRQTASIFKQPVTKVTNHPSNKVKSDPQRM 125
             | |        :|| :   : :      :|:         |:|       | :|||
Db     124 SLSKRPMGRVLNPLREMGVQVEAADGDRMPLTLIGPKTANPITYRVPMASAQVKS----- 178

Qy     126 NEQPRQLFWEKRLQGLSASDVT---EQIIKTMELPKGLQGVGPGSNDETLLSAVASALHT 182
                       | ||:   ||   | ::      | ||| |     ||    |     |
Db     179 ---------AVLLAGLNTPGVTTVIEPVMTRDHTEKMLQGFGADLTVETDKDGVRHIRIT 229

Qy     183 SSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEALMADI 242
           : || :  |  :|:   :|: ||  | :|   |:  :   :   |      |:  :
Db     230 GQGKLVGQ-TIDVPGDPS---STAFPLVAALLVEGSDVTIRNVLMNPTR----TGLILTL 281

Qy     243 LSRAADTEEMDIEMDSGDE 261
             || | ::  :  |::
```

```
Db    282 QEMGADIEVLNARLAGGED 300


RESULT 14
US-08-306-063-7
; Sequence 7, Application US/08306063
; Patent No. 5633435
;  GENERAL INFORMATION:
;    APPLICANT:  Barry, Gerard F.
;    APPLICANT:  Kishore, Ganesh M.
;    APPLICANT:  Padgette, Stephen R.
;    APPLICANT:  Stallings, William C.
;    TITLE OF INVENTION:  Glyphosate Tolerant
;    TITLE OF INVENTION:  5-Enolpyruvylshikimate-3-Phosphate Synthases
;    NUMBER OF SEQUENCES:  69
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dennis R. Hoerner, Jr., Monsanto Co. BB4F
;      STREET:  700 Chesterfield Village Parkway
;      CITY:  St. Louis
;      STATE:  Missouri
;      COUNTRY:  USA
;      ZIP:  63198
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/306,063
;      FILING DATE:  13-SEP-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/749,611
;      FILING DATE:  28-AUG-1991
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/576,537
;      FILING DATE:  31-AUG-1990
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hoerner Jr., Dennis R.
;      REGISTRATION NUMBER:  30,914
;      REFERENCE/DOCKET NUMBER:  38-21(10660)A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (314)537-6099
;      TELEFAX:  (314)537-6047
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  449 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-306-063-7

  Query Match             6.7%;  Score 90.5;  DB 1;  Length 449;
  Best Local Similarity   22.0%;  Pred. No. 1.3;
  Matches   57;  Conservative   38;  Mismatches  123;  Indels   41;  Gaps    9;

Qy      16 IRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGN-------TVDL-SSFDFRTGKMMPS 67
           ||| |      ||:  :  |     :|: |   ||       |: |  ::| :|  :  :
Db      70 IRKEG------DVWIINGVGNGCLLQPEAALDFGNAGTGARLTMGLVGTYDMKTSFIGDA 123

Qy      68 KLQKNKQRLRNDPLNQNKGKPDL--NTTLPIRQTASIFKQPVTKVTNHPSNKVKSDPQRM 125
            | |      :|| :   : :      :|:         |:|      | :|||
Db     124 SLSKRPMGRVLNPLREMGVQVEAADGDRMPLTLIGPKTANPITYRVPMASAQVKS----- 178

Qy     126 NEQPRQLFWEKRLQGLSASDVT---EQIIKTMELPKGLQGVGPGSNDETLLSAVASALHT 182
                   | ||:   ||   | ::      | ||| |     ||    |     |
Db     179 ---------AVLLAGLNTPGVTTVIEPVMTRDHTEKMLQGFGADLTVETDKDGVRHIRIT 229

Qy     183 SSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEALMADI 242
           : || :  |  :|:   :|: ||  | :|   |: :   :   |      |: :
Db     230 GQGKLVGQ-TIDVPGDPS---STAFPLVAALLVEGSDVTIRNVLMNPTR----TGLILTL 281


Qy     243 LSRAADTEEMDIEMDSGDE 261
           || | ::  :  |::
Db     282 QEMGADIEVLNARLAGGED 300


RESULT 15
US-08-833-485-5
; Sequence 5, Application US/08833485
; Patent No. 5804425
;  GENERAL INFORMATION:
;    APPLICANT:  Barry, Gerard F.
;    APPLICANT:  Kishore, Ganesh M.
;    APPLICANT:  Padgette, Stephen R.
;    APPLICANT:  Stallings, William C.
;    TITLE OF INVENTION:  Glyphosate Tolerant
;    TITLE OF INVENTION:  5-Enolpyruvylshikimate-3-Phosphate Synthases
;    NUMBER OF SEQUENCES:  69
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dennis R. Hoerner, Jr., Monsanto Co. BB4F
;      STREET:  700 Chesterfield Village Parkway
;      CITY:  St. Louis
;      STATE:  Missouri
;      COUNTRY:  USA
;      ZIP:  63198
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/833,485
;      FILING DATE:  07-APR-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/306,063
;      FILING DATE:  13-SEP-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/749,611
;      FILING DATE:  28-AUG-1991
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/576,537
;      FILING DATE:  31-AUG-1990
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hoerner Jr., Dennis R.
;      REGISTRATION NUMBER:  30,914
;      REFERENCE/DOCKET NUMBER:  38-21(15117)A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (314)737-6099
;      TELEFAX:  (314)737-6047
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  449 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-833-485-5

  Query Match             6.7%;  Score 90.5;  DB 1;  Length 449;
  Best Local Similarity   22.0%;  Pred. No. 1.3;
  Matches   57;  Conservative   38;  Mismatches  123;  Indels   41;  Gaps    9;

Qy      16 IRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGN-------TVDL-SSFDFRTGKMMPS 67
           ||| |      ||:  :  |     :|: |   ||       |: |  ::| :|  :  :
Db      70 IRKEG------DVWIINGVGNGCLLQPEAALDFGNAGTGARLTMGLVGTYDMKTSFIGDA 123

Qy      68 KLQKNKQRLRNDPLNQNKGKPDL--NTTLPIRQTASIFKQPVTKVTNHPSNKVKSDPQRM 125
            | |      :|| :   : :      :|:         |:|      | :|||
Db     124 SLSKRPMGRVLNPLREMGVQVEAADGDRMPLTLIGPKTANPITYRVPMASAQVKS----- 178
```

```
Qy     126 NEQPRQLFWEKRLQGLSASDVT---EQIIKTMELPKGLQGVGPGSNDETLLSAVASALHT 182
Db     179 ---------AVLLAGLNTPGVTTVIEPVMTRDHTEKMLQGFGADLTVETDKDGVRHIRIT 229

Qy     183 SSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEALMADI 242
Db     230 GQGKLVGQ-TIDVPGDPS---STAFPLVAALLVEGSDVTIRNVLMNPTR----TGLILTL 281

Qy     243 LSRAADTEEMDIEMDSGDE 261
Db     282 QEMGADIEVLNARLAGGED 300
```

Search completed: March 12, 2003, 09:15:37
Job time : 17.4577 secs

GenCore version 5.1.4_p5_4578

OM protein - protein search, using sw model

Run on: March 12, 2003, 03:17:33 ; Search time 36.2051 Seconds
(without alignments)
1491.072 Million cell updates/sec

Title: US-09-554-414B-2_COPY_150_411
Perfect score: 1344
Sequence: 1 MDCPALPPGWKKEEVIRKSG.....LSRAADTEEMDIEMDSGDEA 262

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1344	100.0	411	4	Q9UBB5	Q9ubb5 homo sapien
2	1327	98.7	414	11	Q9Z2E1	Q9z2e1 mus musculu
3	1050	78.1	282	13	Q9PUM9	Q9pum9 xenopus lae
4	1047.5	77.9	303	13	Q9PUM8	Q9pum8 xenopus lae
5	1021	76.0	291	4	O95983	O95983 homo sapien
6	1013	75.4	200	4	O60535	O60535 homo sapien
7	1012.5	75.3	285	11	Q9Z2D8	Q9z2d8 mus musculu
8	451	33.6	302	4	Q9UIS8	Q9uis8 homo sapien
9	446	33.2	249	11	Q9Z2D9	Q9z2d9 mus musculu
10	436.5	32.5	314	5	Q9VHB7	Q9vhb7 drosophila
11	409	30.4	146	4	O95242	O95242 homo sapien
12	384.5	28.6	226	5	Q9V424	Q9v424 drosophila
13	339	25.2	194	4	Q8WWY6	Q8wwy6 homo sapien
14	246	18.3	186	11	Q9D2S3	Q9d2s3 mus musculu
15	246	18.3	186	11	Q9D9H3	Q9d9h3 mus musculu
16	237.5	17.7	588	11	Q9DC19	Q9dc19 mus musculu
17	230.5	17.2	636	11	Q9Z2E2	Q9z2e2 mus musculu
18	214.5	16.0	277	4	Q96EC1	Q96ec1 homo sapien
19	214.5	16.0	503	4	Q9UNZ6	Q9unz6 homo sapien
20	214.5	16.0	549	4	Q9UNZ7	Q9unz7 homo sapien
21	214.5	16.0	556	4	O15248	O15248 homo sapien
22	214.5	16.0	586	4	Q9UNZ8	Q9unz8 homo sapien
23	214.5	16.0	605	4	Q9UNZ9	Q9unz9 homo sapien
24	214.5	16.0	605	4	Q9UIS9	Q9uis9 homo sapien
25	177	13.2	486	6	Q95LG8	Q95lg8 macaca fasc
26	167	12.4	467	13	Q9YGC6	Q9ygc6 xenopus lae
27	164.5	12.2	344	13	O42403	O42403 gallus gall
28	135.5	10.1	554	11	Q8R3R3	Q8r3r3 mus musculu
29	130.5	9.7	554	11	Q9Z2D7	Q9z2d7 mus musculu
30	126	9.4	225	10	Q9LTJ1	Q9ltj1 arabidopsis
31	126	9.4	574	4	Q96F09	Q96f09 homo sapien
32	126	9.4	580	4	O95243	O95243 homo sapien
33	125.5	9.3	155	10	Q9FZP6	Q9fzp6 arabidopsis
34	119	8.9	1627	5	Q962Q0	Q962q0 giardia lam
35	118	8.8	186	10	Q9LYB9	Q9lyb9 arabidopsis
36	117.5	8.7	384	10	Q9XI36	Q9xi36 arabidopsis
37	115	8.6	270	10	Q9FZP7	Q9fzp7 arabidopsis
38	115	8.6	353	10	Q9SLR5	Q9slr5 triticum ae
39	115	8.6	820	10	Q9STZ6	Q9stz6 arabidopsis
40	111	8.3	2003	11	Q91Y51	Q91y51 mus musculu
41	111	8.3	2025	11	Q9R0L6	Q9r0l6 mus musculu
42	110.5	8.2	602	13	Q8QHJ3	Q8qhj3 brachydanio
43	109.5	8.1	182	10	Q9SNC0	Q9snc0 arabidopsis
44	108.5	8.1	865	5	Q9N4L7	Q9n4l7 caenorhabdi
45	108.5	8.1	1972	4	Q9UIF8	Q9uif8 homo sapien

ALIGNMENTS

RESULT 1
Q9UBB5
ID Q9UBB5 PRELIMINARY; PRT; 411 AA.
AC Q9UBB5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Methyl-CpG binding protein 2.
GN MBD2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99373255; PubMed=10441743;
RA Hendrich B., Abbott C., McQueen H., Chambers D., Cross S., Bird A.;
RT "Genomic structure and chromosomal mapping of the murine and human
RT mbd1, mbd2, mbd3, and mbd4 genes.";
RL Mamm. Genome 10:906-912(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=98449942; PubMed=9774669;
RA Hendrich B., Bird A.;
RT "Identification and characterization of a family of mammalian methyl-
RT CpG binding proteins.";
RL Mol. Cell. Biol. 18:6538-6547(1998).
DR EMBL; AF120993; AAD56597.1; -.
DR EMBL; AF120988; AAD56597.1; JOINED.
DR EMBL; AF120989; AAD56597.1; JOINED.
DR EMBL; AF120990; AAD56597.1; JOINED.
DR EMBL; AF120991; AAD56597.1; JOINED.
DR EMBL; AF120992; AAD56597.1; JOINED.
DR EMBL; AF072242; AAC68871.1; -.
DR InterPro; IPR001739; Methyl-CpG_bind.
DR Pfam; PF01429; MBD; 1.
DR SMART; SM00391; MBD; 1.
SQ SEQUENCE 411 AA; 43254 MW; FC4E5E0CF9BA0FFA CRC64;

```
  Query Match           100.0%;  Score 1344;  DB 4;  Length 411;
  Best Local Similarity  100.0%;  Pred. No. 7.8e-101;
  Matches  262;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MDCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNTVDLSSFDFR 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     150 MDCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNTVDLSSFDFR 209

Qy      61 TGKMMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKVKS 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     210 TGKMMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKVKS 269

Qy     121 DPQRMNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSAVASAL 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     270 DPQRMNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSAVASAL 329

Qy     181 HTSSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEALMA 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     330 HTSSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEALMA 389

Qy     241 DILSRAADTEEMDIEMDSGDEA 262
           ||||||||||||||||||||||
Db     390 DILSRAADTEEMDIEMDSGDEA 411
```

RESULT 2
Q9Z2E1
ID Q9Z2E1 PRELIMINARY; PRT; 414 AA.
AC Q9Z2E1;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Methyl-CpG binding protein MBD2.
GN MBD2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BLACK/6;
RX MEDLINE=98449942; PubMed=9774669;
RA Hendrich B., Bird A.;
RT "Identification and characterization of a family of mammalian methyl-
RT CpG binding proteins.";
RL Mol. Cell. Biol. 18:6538-6547(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=129;
RX MEDLINE=99373255; PubMed=10441743;
RA Hendrich B., Abbott C., McQueen H., Chambers D., Cross S., Bird A.;
RT "Genomic structure and chromosomal mapping of the murine and human
RT mbd1, mbd2, mbd3, and mbd4 genes.";
RL Mamm. Genome 10:906-912(1999).
DR EMBL; AF072243; AAC68872.1; -.
DR EMBL; AF120986; AAD50372.1; -.
DR EMBL; AF120983; AAD50372.1; JOINED.
DR EMBL; AF120984; AAD50372.1; JOINED.
DR EMBL; AF120985; AAD50372.1; JOINED.
DR MGD; MGI:1333813; Mbd2.
DR InterPro; IPR001739; Methyl-CpG_bind.
DR Pfam; PF01429; MBD; 1.
DR SMART; SM00391; MBD; 1.
SQ SEQUENCE 414 AA; 43543 MW; 9601D95E347E8E53 CRC64;

```
  Query Match           98.7%;  Score 1327;  DB 11;  Length 414;
  Best Local Similarity  98.5%;  Pred. No. 1.9e-99;
  Matches  258;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 MDCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNTVDLSSFDFR 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     153 MDCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNAVDLSSFDFR 212

Qy      61 TGKMMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKVKS 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     213 TGKMMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTASIFKQPVTKFTNHPSNKVKS 272

Qy     121 DPQRMNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSAVASAL 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     273 DPQRMNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSAVASAL 332

Qy     181 HTSSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEALMA 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     333 HTSSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEALMA 392

Qy     241 DILSRAADTEEMDIEMDSGDEA 262
           ||||||||||:||:||||||||
Db     393 DILSRAADTEEVDIDMDSGDEA 414
```

RESULT 3
Q9PUM9
ID Q9PUM9 PRELIMINARY; PRT; 282 AA.
AC Q9PUM9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Methyl-CpG binding protein MBD3.
GN MBD3.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA Wade P.A., Gegonne A., Jones P.L., Ballestar E., Aubry F.,
RA Wolffe A.P.;
RT "The Mi-2 histone deacetylase complex couples DNA methylation to
RT chromatin remodeling and histone deacetylation.";
RL Nat. Genet. 0:0-0(1999).
DR EMBL; AF170346; AAD55389.1; -.
DR InterPro; IPR001739; Methyl-CpG_bind.
DR Pfam; PF01429; MBD; 1.
DR SMART; SM00391; MBD; 1.
SQ SEQUENCE 282 AA; 31692 MW; 6891EFA05B0371E9 CRC64;

```
  Query Match           78.1%;  Score 1050;  DB 13;  Length 282;
  Best Local Similarity  75.5%;  Pred. No. 3.2e-77;
  Matches  200;  Conservative   33;  Mismatches   26;  Indels    6;  Gaps    2;

Qy       2 DCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNTVDLSSFDFRT 61
           :| ||  ||||||| |:||||||||||:||||||||||||||||||::|||:|||||
Db       7 ECSAL-QGWKKEEVTRRSGLSAGKSDVYYYSPSGKKFRSKPQLARYLGNSMDLSTFDFRT 65

Qy      62 GKMMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKVKSD 121
           |||: ||: ||:||:| | |||:|||||||| ||:||||||||||||||||:||||||
Db      66 GKMLMSKINKNRQRMRYDGLNQSKGKPDLNTALPVRQTASIFKQPVTKVTNHPTNKVKSD 125

Qy     122 PQRMNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSAVASALH 181
           ||:  :|||||||||:| ||:| |: |:::|||||||||||||||  |||||||:|||||
Db     126 PQKAVDQPRQLFWEKKLSGLNAFDIAEELVKTMELPKGLQGVGPGCTDETLLSAIASALH 185

Qy     182 TSSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEALMAD 241
           ||: |||||:|||||||| ||||||||||||:|||||||||| ||||||||||||||||
Db     186 TSTMPITGQLSAAVEKNPGVWLNTSQPLCKAFMVTDEDIRKQEELVQQVRKKLEEALMAD 245

Qy     242 ILSRAADTEE-----MDIEMDSGDE 261
           :|:   :  :     :| ::|  :|
Db     246 MLAHVEEISKDGGAPLDKDIDDEEE 270
```

RESULT 4
Q9PUM8
ID Q9PUM8 PRELIMINARY; PRT; 303 AA.
AC Q9PUM8;

```
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Methyl-CpG binding protein MBD3 long form.
GN   MBD3.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Wade P.A., Gegonne A., Jones P.L., Ballestar E., Aubry F.,
RA   Wolffe A.P.;
RT   "The Mi-2 histone deacetylase complex couples DNA methylation to
RT   chromatin remodeling and histone deacetylation.";
RL   Nat. Genet. 0:0-0(1999).
DR   EMBL; AF170347; AAD55390.1; -.
DR   InterPro; IPR001739; Methyl-CpG_bind.
DR   Pfam; PF01429; MBD; 1.
DR   SMART; SM00391; MBD; 1.
SQ   SEQUENCE   303 AA;  34078 MW;  864B727E50EF710B CRC64;

  Query Match             77.9%;  Score 1047.5;  DB 13;  Length 303;
  Best Local Similarity   70.5%;  Pred. No. 5.6e-77;
  Matches  201;  Conservative   33;  Mismatches   26;  Indels   25;  Gaps    2;

Qy       2 DCPALPPGWKKEEVIRKSGLSAGKSDVYYFS-------------------PSGKKFRSK 41
           :| ||| ||||||| |:|||||||||||||:|                   |||||||||
Db       7 ECSALPQGWKKEEVTRRSGLSAGKSDVYYYSSSPSRYNRSLRDRVGCLNINPSGKKFRSK 66

Qy      42 PQLARYLGNTVDLSSFDFRTGKMMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTAS 101
           |||||||||::|||:||||||||: ||: ||:||:| | |||:|||||||| ||:|||||
Db      67 PQLARYLGNSMDLSTFDFRTGKMLMSKINKNRQRMRYDGLNQSKGKPDLNTALPVRQTAS 126

Qy     102 IFKQPVTKVTNHPSNKVKSDPQRMNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQ 161
           |||||||||||||:||||||||:  :|||||||||:| ||:| |: |:::|||||||||||
Db     127 IFKQPVTKVTNHPTNKVKSDPQKAVDQPRQLFWEKKLSGLNAFDIAEELVKTMELPKGLQ 186

Qy     162 GVGPGSNDETLLSAVASALHTSSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIR 221
           |||||  ||||||:|||||||: |||||:|||||||| ||||||||||||||:|||||||
Db     187 GVGPGCTDETLLSAIASALHTSTMPITGQLSAAVEKNPGVWLNTSQPLCKAFMVTDEDIR 246

Qy     222 KQEERVQQVRKKLEEALMADILSRAADTEE-----MDIEMDSGDE 261
           |||| |||||||||||||||:|:   :  :     :| ::|  :|
Db     247 KQEELVQQVRKKLEEALMADMLAHVEEISKDGGAPLDKDIDDEEE 291


RESULT 5
O95983
ID   O95983      PRELIMINARY;   PRT;   291 AA.
AC   O95983;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Methyl-CpG binding protein MBD3.
GN   MBD3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Lamerdin J.E., McCready P.M., Skowronski E., Viswanathan V.,
RA   Burkhart-Schultz K.J., Gordon L., Dias J., Ramirez M., Stilwagen S.,
RA   Phan H., Velasco N., Do L., Regala W., Terry A., Garnes J.,
RA   Danganan L., Erler A., Christensen M., Georgescu A., Avila J., Liu S.,
RA   Attix C., Andreise T., Trankheim M., Amico-Keller G., Coefield J.,
RA   Duarte S., Lucas S., Bruce R., Thomas P., Quan G., Kronmiller B.,
RA   Arellano A., Saunders C., Ow D., Nolan M., Trong S., Kobayashi A.,
RA   Olsen A.S., Carrano A.V.;
RT   "Sequence analysis of a 3.5 Mb contig in human 19p13.3 containing a
RT   serine protease gene cluster.";
RL   Submitted (NOV-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98449942; PubMed=9774669;
RA   Hendrich B., Bird A.;
RT   "Identification and characterization of a family of mammalian methyl-
RT   CpG binding proteins.";
RL   Mol. Cell. Biol. 18:6538-6547(1998).
DR   EMBL; AC005943; AAC72104.1; -.
DR   EMBL; AF072247; AAC68876.1; -.
DR   InterPro; IPR001739; Methyl-CpG_bind.
DR   Pfam; PF01429; MBD; 1.
DR   SMART; SM00391; MBD; 1.
SQ   SEQUENCE   291 AA;  32844 MW;  B62134DD1BEB636B CRC64;

  Query Match             76.0%;  Score 1021;  DB 4;  Length 291;
  Best Local Similarity   72.1%;  Pred. No. 7.5e-75;
  Matches  196;  Conservative   34;  Mismatches   30;  Indels   12;  Gaps    2;

Qy       2 DCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNTVDLSSFDFRT 61
           :||||| ||::||| |:|||||  ||:|:||||||||||||||| ::|||:|||||
Db       7 ECPALPQGWEREEVPRRSGLSAGHRDVFYYSPSGKKFRSKPQLARYLGGSMDLSTFDFRT 66

Qy      62 GKMMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKVKSD 121
           |||: ||: |::||:| |  || |||||||| ||:|||||||||||:||||||||||
Db      67 GKMLMSKMNKSRQRVRYDSSNQVKGKPDLNTALPVRQTASIFKQPVTKITNHPSNKVKSD 126

Qy     122 PQRMNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSAVASALH 181
           ||:  :|||||||||:| ||:| |: |:::|||:|||||||||  |||||||:|||||
Db     127 PQKAVDQPRQLFWEKKLSGLNAFDIAEELVKTMDLPKGLQGVGPGCTDETLLSAIASALH 186

Qy     182 TSSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEALMAD 241
           ||: |||||:|||||||| |||||:|||||||:|||||||||| ||||||:||||||||
Db     187 TSTMPITGQLSAAVEKNPGVWLNTTQPLCKAFMVTDEDIRKQEELVQQVRKRLEEALMAD 246

Qy     242 ILSR----AADTE--------EMDIEMDSGDE 261
           :|:       | | |       | | | |  :|
Db     247 MLAHVEELARDGEAPLDKACAEDDDEEDEEEE 278


RESULT 6
O60535
ID   O60535      PRELIMINARY;      PRT;   200 AA.
AC   O60535;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Antigen NY-CO-41 (Fragment).
GN   NY-CO-41.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=COLON CANCER METASTASIS TO LUNG;
RX   MEDLINE=98272252; PubMed=9610721;
RA   Scanlan M.J., Chen Y.T., Williamson B., Gure A.O., Stockert E.,
RA   Gordan J.D., Tureci O., Sahin U., Pfreundschuh M., Old L.J.;
RT   "Characterization of human colon cancer antigens recognized by
RT   autologous antibodies.";
RL   Int. J. Cancer 76:652-658(1998).
DR   EMBL; AF039701; AAC18050.1; -.
FT   NON_TER       1       1
SQ   SEQUENCE   200 AA;  22270 MW;  A70B46B35CA55AFD CRC64;

  Query Match             75.4%;  Score 1013;  DB 4;  Length 200;
  Best Local Similarity   100.0%;  Pred. No. 2e-74;
  Matches  200;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      63 KMMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKVKSDP 122
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 KMMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKVKSDP 60
```

```
Qy    123 QRMNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSAVASALHT 182
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 QRMNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSAVASALHT 120

Qy    183 SSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEALMADI 242
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 SSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEALMADI 180

Qy    243 LSRAADTEEMDIEMDSGDEA 262
          ||||||||||||||||||||
Db    181 LSRAADTEEMDIEMDSGDEA 200


RESULT 7
Q9Z2D8
ID   Q9Z2D8      PRELIMINARY;      PRT;   285 AA.
AC   Q9Z2D8;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Methyl-CpG binding  protein MBD3.
GN   MBD3.
OS   Mus musculus (Mouse), and
OS   Mus musculus domesticus (western European house mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090, 10092;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98449942; PubMed=9774669;
RA   Hendrich B., Bird A.;
RT   "Identification and characterization of a family of mammalian methyl-
RT   CpG binding proteins.";
RL   Mol. Cell. Biol. 18:6538-6547(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129;
RX   MEDLINE=99373255; PubMed=10441743;
RA   Hendrich B., Abbott C., McQueen H., Chambers D., Cross S., Bird A.;
RT   "Genomic structure and chromosomal mapping of the murine and human
RT   mbd1, mbd2, mbd3, and mbd4 genes.";
RL   Mamm. Genome 10:906-912(1999).
DR   EMBL; AF072248; AAC68877.1; -.
DR   EMBL; AF120995; AAD48909.1; -.
DR   MGD; MGI:1333812; Mbd3.
DR   InterPro; IPR001739; Methyl-CpG_bind.
DR   Pfam; PF01429; MBD; 1.
DR   SMART; SM00391; MBD; 1.
SQ   SEQUENCE   285 AA;  32168 MW;  E4E57BD48463643F CRC64;

  Query Match             75.3%;  Score 1012.5;  DB 11;  Length 285;
  Best Local Similarity   75.3%;  Pred. No. 3.6e-74;
  Matches  189;  Conservative   35;  Mismatches   24;  Indels    3;  Gaps    1;

Qy      2 DCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNTVDLSSFDFRT 61
          :||||| ||::||| |:|||||  ||:|:||||||||||||||| ::|||:|||||
Db      7 ECPALPQGWEREEVPRRSGLSAGHRDVFYYSPSGKKFRSKPQLARYLGGSMDLSTFDFRT 66

Qy     62 GKMMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKVKSD 121
          |||: :|: |::||:| |  || |||||||| ||:|||||||||||||:||||||||||
Db     67 GKMLMNKMNKSRQRVRYDSSNQVKGKPDLNTALPVRQTASIFKQPVTKITNHPSNKVKSD 126

Qy    122 PQRMNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSAVASALH 181
          ||:  :|||||||||:| |||| |: |::::||:||||||||||  |||||||:|||||
Db    127 PQKAVDQPRQLFWEKKLSGLSAFDIAEELVRTMDLPKGLQGVGPGCTDETLLSAIASALH 186

Qy    182 TSSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEALMAD 241
          ||: |||||:|||||||| |||||:|||||||:|||:|||||||  |||||:|||||||||
Db    187 TSTLPITGQLSAAVEKNPGVWLNTAQPLCKAFMVTDDDIRKQEELVQQVRKRLEEALMAD 246

Qy    242 ILSRAADTEEM 252
          ::|   |  ||:
```

```
Db    247 ML---AHVEEL 254


RESULT 8
Q9UIS8
ID   Q9UIS8      PRELIMINARY;      PRT;   302 AA.
AC   Q9UIS8;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Testis-specific methyl-CpG binding protein 2.
GN   MBD2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99373255; PubMed=10441743;
RA   Hendrich B., Abbott C., McQueen H., Chambers D., Cross S., Bird A.;
RT   "Genomic structure and chromosomal mapping of the murine and human
RT   Mbd1, Mbd2, Mbd3, and Mbd4 genes.";
RL   Mamm. Genome 10:906-912(1999).
DR   EMBL; AF120989; AAD56596.1; -.
DR   EMBL; AF120988; AAD56596.1; JOINED.
DR   InterPro; IPR001739; Methyl-CpG_bind.
DR   Pfam; PF01429; MBD; 1.
DR   SMART; SM00391; MBD; 1.
SQ   SEQUENCE   302 AA;  31745 MW;  CCEC65D926222717 CRC64;

  Query Match             33.6%;  Score 451;  DB 4;  Length 302;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-28;
  Matches   85;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 MDCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNTVDLSSFDFR 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    150 MDCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNTVDLSSFDFR 209

Qy     61 TGKMMPSKLQKNKQRLRNDPLNQNK 85
          |||||||||||||||||||||||||
Db    210 TGKMMPSKLQKNKQRLRNDPLNQNK 234


RESULT 9
Q9Z2D9
ID   Q9Z2D9      PRELIMINARY;      PRT;   249 AA.
AC   Q9Z2D9;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Testis specific methyl-CpG binding protein MBD2.
GN   MBD2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BLACK/6; TISSUE=TESTIS;
RX   MEDLINE=98449942; PubMed=9774669;
RA   Hendrich B., Bird A.;
RT   "Identification and characterization of a family of mammalian methyl-
RT   CpG binding proteins.";
RL   Mol. Cell. Biol. 18:6538-6547(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129;
RX   MEDLINE=99373255; PubMed=10441743;
RA   Hendrich B., Abbott C., McQueen H., Chambers D., Cross S., Bird A.;
RT   "Genomic structure and chromosomal mapping of the murine and human
RT   mbd1, mbd2, mbd3, and mbd4 genes.";
RL   Mamm. Genome 10:906-912(1999).
DR   EMBL; AF072245; AAC68874.1; -.
```

```
DR   EMBL; AF120983; AAD50373.1; -.
DR   MGD; MGI:1333813; Mbd2.
DR   InterPro; IPR001739; Methyl-CpG_bind.
DR   Pfam; PF01429; MBD; 1.
DR   SMART; SM00391; MBD; 1.
SQ   SEQUENCE   249 AA;  25493 MW;  530D751E8E5C9540 CRC64;

  Query Match             33.2%;  Score 446;  DB 11;  Length 249;
  Best Local Similarity   98.8%;  Pred. No. 2.5e-28;
  Matches   84;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 MDCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNTVDLSSFDFR 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db     153 MDCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNAVDLSSFDFR 212

Qy      61 TGKMMPSKLQKNKQRLRNDPLNQNK 85
           |||||||||||||||||||||||||
Db     213 TGKMMPSKLQKNKQRLRNDPLNQNK 237


RESULT 10
Q9VHB7
ID   Q9VHB7      PRELIMINARY;      PRT;   314 AA.
AC   Q9VHB7;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Methyl-CpG-binding-domain-like-protein.
GN   METHYL-CPG-BINDING-DOMAIN-LIKE-PROTEIN OR CG8208.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003683; AAF54400.1; -.
DR   FlyBase; FBgn0027950; methyl-CpG-binding-domain-like-protein.
DR   InterPro; IPR001739; Methyl-CpG_bind.
DR   Pfam; PF01429; MBD; 1.
DR   SMART; SM00391; MBD; 1.
SQ   SEQUENCE   314 AA;  33803 MW;  7B6BDFE63873D230 CRC64;

  Query Match             32.5%;  Score 436.5;  DB 5;  Length 314;
  Best Local Similarity   35.0%;  Pred. No. 2e-27;
  Matches  108;  Conservative   48;  Mismatches   72;  Indels   81;  Gaps    9;

Qy       1 MDCPALPPGWKKEEVIRKSGLSA-------------------GKSDVYYFSPSGKKFRS 40
             :||  || ||:::|| |||| ||                     | ||:|:||:||:
Db      14 VDCSVLPKGWQRDEV-RKSGSSANNNASSNNNSSATASSNNNNNKVDVFYYSPTGKRAEG 72

Qy      41 KPQLARYLGNTVDLSSFDFRTGKM-----------------MP--------------SKL 69
           |||         |::  ||: |||                 ||              |
Db      73 KPQ---------DIAIPDFQPGKMPHCALPSPSISLYRCSAMPLPIASGGGNGATSGSAA 123

Qy      70 QKNKQRLRNDPLNQNKG---------------KPDLNTTLPIRQTASIFKQPVTKVTNHP 114
              |::      |                  : |::   ||||||||||||| : ||
Db     124 NALKRKFARSQGGNAAGAAGAAPPAATASRALRTDVSLVPPIRQTASIFKQPVTVIRNHK 183

Qy     115 SN--KVKSDPQR-MNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDET 171
             :  | |::|:    |:|:|||||||: | |   : : :  : ||| :: |||  |::|
Db     184 QDPAKAKNEPKHGTREKPKQLFWEKRLERLRACHDSGEELDDISLPKTIRTVGPNVNEQT 243

Qy     172 LLSAVASALHTSSAPITGQVS--AAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQ 229
           :| :||:|||  :| : || |  | : ||   ::|  |||  | |::::|||||:||
Db     244 VLQSVATALHMLNAGVHGQSSTKADLTKNAMAFMNPEQPLMHAVIISEDDIRKQEDRVGV 303

Qy     230 VRKKLEEAL 238
             |:||::||
Db     304 ARRKLQDAL 312


RESULT 11
O95242
ID   O95242      PRELIMINARY;      PRT;   146 AA.
AC   O95242;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Testis specific methyl-CpG binding protein MBD2 (Fragment).
GN   MBD2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=TESTIS;
RX   MEDLINE=98449942; PubMed=9774669;
RA   Hendrich B., Bird A.;
RT   "Identification and characterization of a family of mammalian methyl-
RT   CpG binding proteins.";
RL   Mol. Cell. Biol. 18:6538-6547(1998).
DR   EMBL; AF072246; AAC68875.1; -.
DR   InterPro; IPR001739; Methyl-CpG_bind.
DR   Pfam; PF01429; MBD; 1.
DR   SMART; SM00391; MBD; 1.
FT   NON_TER       1      1
SQ   SEQUENCE   146 AA;  16931 MW;  9D6CC3CFE140EEEA CRC64;

  Query Match             30.4%;  Score 409;  DB 4;  Length 146;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-25;
  Matches   78;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8 PGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNTVDLSSFDFRTGKMMPS 67
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      1 PGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNTVDLSSFDFRTGKMMPS 60

Qy     68 KLQKNKQRLRNDPLNQNK 85
          ||||||||||||||||||
Db     61 KLQKNKQRLRNDPLNQNK 78
```

RESULT 12
Q9V424
ID   Q9V424   PRELIMINARY;   PRT;   226 AA.
AC   Q9V424;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Methyl-CpG-binding-domain-like-protein.
GN   METHYL-CPG-BINDING-DOMAIN-LIKE-PROTEIN OR CG8208.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Wade P.A., Gegonne A., Jones P.L., Ballestar E., Aubry F.,
RA   Wolffe A.P.;
RT   "The Mi-2 histone deacetylase complex couples DNA methylation to
RT   chromatin remodeling and histone deacetylation.";
RL   Nat. Genet. 0:0-0(1999).
DR   EMBL; AE003683; AAF54401.1; -.
DR   EMBL; AF171098; AAD55391.1; -.
DR   FlyBase; FBgn0027950; methyl-CpG-binding-domain-like-protein.
DR   InterPro; IPR001739; Methyl-CpG_bind.
DR   Pfam; PF01429; MBD; 1.
DR   SMART; SM00391; MBD; 1.
SQ   SEQUENCE   226 AA; 25292 MW; 26F2DD296442B4D3 CRC64;

```
  Query Match             28.6%;  Score 384.5;  DB 5;  Length 226;
  Best Local Similarity   38.3%;  Pred. No. 2.1e-23;
  Matches   93;  Conservative   43;  Mismatches   70;  Indels   37;  Gaps    7;

Qy      1 MDCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNTVDLSSFDFR 60
          :||  || ||:::|| |||| ||        : |     |    :    | ||:  :
Db     14 VDCSVLPKGWQRDEV-RKSGSSANN------NASSNNNSSATASSNNNNNKVDVFYY--- 63

Qy     61 TGKMMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSN--KV 118
                  : ||           |::   |||||||||||||| : ||  :  |
Db     64 -----------SRALRT----------DVSLVPPIRQTASIFKQPVTVIRNHKQDPAKA 101

Qy    119 KSDPQR-MNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSAVA 177
          |::|:    |:|:|||||||||: | |   : : :  : ||| :: |||  |::|:| :||
Db    102 KNEPKHGTREKPKQLFWEKRLERLRACHDSGEELDDISLPKTIRTVGPNVNEQTVLQSVA 161

Qy    178 SALHTSSAPITGQVS--AAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLE 235
          :|||  :| : || |  | : ||   ::|  |||  | |::::|||||:||   |:||:
Db    162 TALHMLNAGVHGQSSTKADLTKNAMAFMNPEQPLMHAVIISEDDIRKQEDRVGVARRKLQ 221

Qy    236 EAL 238
          :||
Db    222 DAL 224
```

RESULT 13
Q8WWY6
ID   Q8WWY6   PRELIMINARY;   PRT;   194 AA.
AC   Q8WWY6;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Methyl-CpG binding domain protein 3-like.
GN   MBD3L.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Jiang C.L., Szabo P.E., O'Connor T.R., Mann J.R., Pfeifer G.P.;
RT   "MBD3L, a gene homologous to the methyl-CpG-binding domain protein
RT   genes MBD3 and MBD2, is expressed specifically in postmeiotic cells of
RT   the testis.";
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY038022; AAK72591.1; -.
SQ   SEQUENCE   194 AA; 21562 MW; F9EC132569145DC9 CRC64;

```
  Query Match             25.2%;  Score 339;  DB 4;  Length 194;
  Best Local Similarity   41.8%;  Pred. No. 8.4e-20;
  Matches   76;  Conservative   39;  Mismatches   65;  Indels    2;  Gaps    2;

Qy     69 LQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKVKSDP-QRMNE 127
          : |: || : | :|| | || |:|::|:| ::  ||:|||::| || |:|:    :   |
Db      1 MAKSSQRKQRDCVNQCKSKPGLSTSIPLRMSSYTFKRPVTRITPHPGNEVRYHQWEESLE 60

Qy    128 QPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSAVASAL-HTSSAP 186
          :|:|: |::|||||.|     ::  |::|   || : |     :||  :|| | |:   |
Db     61 KPQQVCWQRRLQGLQAYSSAGELSSTLDLANTLQKLVPSYTGGSLLEDLASGLEHSCPMP 120

Qy    187 ITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEALMADILSRA 246
              | |||  ||  :  || ||| |:||:|||||| :|: ||::|  ||:|| |:
Db    121 HLACSSDAVEIIPAEGVGISQLLCKQFLVTEEDIRKQEGKVKTVRERLAIALIADGLANE 180

Qy    247 AD 248
          |:
Db    181 AE 182
```

```
RESULT 14
Q9D2S3
ID   Q9D2S3      PRELIMINARY;      PRT;   186 AA.
AC   Q9D2S3;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   1700095H13Rik protein.
GN   1700095H13RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=TESTIS;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
DR   EMBL; AK018931; BAB31487.1; -.
DR   MGD; MGI:1925722; 1700095H13Rik.
SQ   SEQUENCE   186 AA;  20659 MW;  EEF450ABF69D2AF0 CRC64;

  Query Match             18.3%;  Score 246;  DB 11;  Length 186;
  Best Local Similarity   35.9%;  Pred. No. 2.7e-12;
  Matches   66;  Conservative   36;  Mismatches   68;  Indels   14;  Gaps    5;

Qy      71 KNKQRLRNDPLNQNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKVK-SDPQRMNEQP 129
           |  || : |   :|  || |:|::|:| ::  ||:||||:|:|  |:|:    :   |:|
Db       3 KTSQRKQCD--CENPSKPCLSTSIPLRMSSYTFKRPVTKITSHLGNEVRYYQWEETLEKP 60

Qy     130 RQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSAVASALHT--SSAPI 187
           |  |:|||||| |     ::: | :| | |:       | |    ||||  |  :|  |
Db      61 EQASWQKRLQGLQAYSSAGELLSTSDLAKTLK-------DLTSTDTVASASDTQATSIDI 113

Qy     188 TGQVSAAVEKNPAVWLNTSQP--LCKAFIVTDEDIRKQEERVQQVRKKLEEALMADILSR 245
           |   :     : |  :  : |  ||| |:||::||  || :|:  |::|  ||:|  |:
Db     114 TSVPTLESSSHLANMIPEAGPQILCKEFLVTEQDIINQERKVKIARERLAVALIAHKLAS 173

Qy     246 AADT 249
             :|
Db     174 EMET 177


RESULT 15
Q9D9H3
ID   Q9D9H3      PRELIMINARY;      PRT;   186 AA.
AC   Q9D9H3;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   1700070G05Rik protein (Methyl-CpG binding domain protein 3-like).
GN   1700070G05RIK OR MBD3L.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=TESTIS;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J;
RA   Jiang C.L., Szabo P.E., O'Connor T.R., Mann J.R., Pfeifer G.P.;
RT   "MBD3L, a gene homologous to the methyl-CpG-binding domain protein
RT   genes MBD3 and MBD2, is expressed specifically in postmeiotic cells of
RT   the testis.";
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK006927; BAB24794.1; -.
DR   EMBL; AY038023; AAK72592.1; -.
DR   MGD; MGI:1920753; 1700070G05Rik.
SQ   SEQUENCE   186 AA;  20675 MW;  F5F45BABE69D2AF0 CRC64;

  Query Match             18.3%;  Score 246;  DB 11;  Length 186;
  Best Local Similarity   35.9%;  Pred. No. 2.7e-12;
  Matches   66;  Conservative   36;  Mismatches   68;  Indels   14;  Gaps    5;

Qy      71 KNKQRLRNDPLNQNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKVK-SDPQRMNEQP 129
           |  || : |   :|  || |:|::|:| ::  ||:||||:|:|  |:|:    :   |:|
Db       3 KTSQRKQCD--CENPSKPCLSTSIPLRMSSYTFKRPVTKITSHLGNEVRYYQWEETLEKP 60

Qy     130 RQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSAVASALHT--SSAPI 187
           |  |:|||||| |     ::: | :| | |:       | |    ||||  |  :|  |
Db      61 EQACWQKRLQGLQAYSSAGELLSTSDLAKTLK-------DLTSTDTVASASDTQATSIDI 113

Qy     188 TGQVSAAVEKNPAVWLNTSQP--LCKAFIVTDEDIRKQEERVQQVRKKLEEALMADILSR 245
           |   :     : |  :  : |  ||| |:||::||  || :|:  |::|  ||:|  |:
Db     114 TSVPTLESSSHLANMIPEAGPQILCKEFLVTEQDIINQERKVKIARERLAVALIAHKLAS 173

Qy     246 AADT 249
             :|
Db     174 EMET 177


Search completed: March 12, 2003, 09:12:43
Job time : 39.2051 secs
```

GenCore version 5.1.4_p5_4578

OM protein - protein search, using sw model

Run on: March 12, 2003, 03:17:33 ; Search time 56.7949 Seconds
(without alignments)
1491.072 Million cell updates/sec

Title: US-09-554-414B-2
Perfect score: 2167
Sequence: 1 MRAHPGGGRCCPEQEEGESA.....LSRAADTEEMDIEMDSGDEA 411

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2167	100.0	411	4	Q9UBB5	Q9ubb5 homo sapien
2	2091.5	96.5	414	11	Q9Z2E1	Q9z2e1 mus musculu
3	1274	58.8	302	4	Q9UIS8	Q9uis8 homo sapien
4	1210.5	55.9	249	11	Q9Z2D9	Q9z2d9 mus musculu
5	1059	48.9	282	13	Q9PUM9	Q9pum9 xenopus lae
6	1056.5	48.8	303	13	Q9PUM8	Q9pum8 xenopus lae
7	1030	47.5	291	4	O95983	O95983 homo sapien
8	1021.5	47.1	285	11	Q9Z2D8	Q9z2d8 mus musculu
9	1013	46.7	200	4	O60535	O60535 homo sapien
10	452.5	20.9	314	5	Q9VHB7	Q9vhb7 drosophila
11	409	18.9	146	4	O95242	O95242 homo sapien
12	400.5	18.5	226	5	Q9V424	Q9v424 drosophila
13	339	15.6	194	4	Q8WWY6	Q8wwy6 homo sapien
14	271	12.5	619	12	Q9IPQ9	Q9ipq9 cynomolgus
15	264.5	12.2	588	12	Q9IPQ8	Q9ipq8 cynomolgus
16	246	11.4	186	11	Q9D2S3	Q9d2s3 mus musculu
17	246	11.4	186	11	Q9D9H3	Q9d9h3 mus musculu
18	243.5	11.2	511	12	O91332	O91332 cercopithic
19	237.5	11.0	588	11	Q9DC19	Q9dc19 mus musculu
20	230.5	10.6	636	11	Q9Z2E2	Q9z2e2 mus musculu
21	226	10.4	476	12	Q80890	Q80890 herpesvirus
22	225	10.4	592	16	Q9PF60	Q9pf60 xylella fas
23	220	10.2	221	10	O65514	O65514 arabidopsis
24	218.5	10.1	610	5	Q9V5V8	Q9v5v8 drosophila
25	217.5	10.0	385	5	Q93424	Q93424 caenorhabdi
26	216	10.0	175	10	Q9LSN6	Q9lsn6 arabidopsis
27	214.5	9.9	277	4	Q96EC1	Q96ec1 homo sapien
28	214.5	9.9	503	4	Q9UNZ6	Q9unz6 homo sapien
29	214.5	9.9	549	4	Q9UNZ7	Q9unz7 homo sapien
30	214.5	9.9	556	4	O15248	O15248 homo sapien
31	214.5	9.9	586	4	Q9UNZ8	Q9unz8 homo sapien
32	214.5	9.9	605	4	Q9UNZ9	Q9unz9 homo sapien
33	214.5	9.9	605	4	Q9UIS9	Q9uis9 homo sapien
34	214	9.9	271	10	Q08529	Q08529 nicotiana t
35	211.5	9.8	291	10	Q39337	Q39337 brassica na
36	209	9.6	344	13	O42403	O42403 gallus gall
37	208.5	9.6	396	10	O65450	O65450 arabidopsis
38	207.5	9.6	174	10	Q9LTP5	Q9ltp5 arabidopsis
39	207	9.6	255	10	Q9SIH2	Q9sih2 arabidopsis
40	205.5	9.5	486	6	Q95LG8	Q95lg8 macaca fasc
41	203	9.4	1432	10	Q9FPR8	Q9fpr8 chlamydomon
42	202.5	9.3	302	10	Q9SL09	Q9sl09 arabidopsis
43	202	9.3	293	13	Q9DEX9	Q9dex9 cyprinus ca
44	201	9.3	147	5	Q24348	Q24348 drosophila
45	200.5	9.3	173	10	Q41191	Q41191 arabidopsis

ALIGNMENTS

RESULT 1
Q9UBB5
ID Q9UBB5 PRELIMINARY; PRT; 411 AA.
AC Q9UBB5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Methyl-CpG binding protein 2.
GN MBD2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99373255; PubMed=10441743;
RA Hendrich B., Abbott C., McQueen H., Chambers D., Cross S., Bird A.;
RT "Genomic structure and chromosomal mapping of the murine and human
RT mbd1, mbd2, mbd3, and mbd4 genes.";
RL Mamm. Genome 10:906-912(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=98449942; PubMed=9774669;
RA Hendrich B., Bird A.;
RT "Identification and characterization of a family of mammalian methyl-
RT CpG binding proteins.";
RL Mol. Cell. Biol. 18:6538-6547(1998).
DR EMBL; AF120993; AAD56597.1; -.
DR EMBL; AF120988; AAD56597.1; JOINED.
DR EMBL; AF120989; AAD56597.1; JOINED.
DR EMBL; AF120990; AAD56597.1; JOINED.
DR EMBL; AF120991; AAD56597.1; JOINED.
DR EMBL; AF120992; AAD56597.1; JOINED.
DR EMBL; AF072242; AAC68871.1; -.
DR InterPro; IPR001739; Methyl-CpG_bind.
DR Pfam; PF01429; MBD; 1.
DR SMART; SM00391; MBD; 1.
SQ SEQUENCE 411 AA; 43254 MW; FC4E5E0CF9BA0FFA CRC64;

```
  Query Match           100.0%;  Score 2167;  DB 4;  Length 411;
  Best Local Similarity  100.0%;  Pred. No. 1.2e-152;
  Matches  411;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MRAHPGGGRCCPEQEEGESAAGGSGAGGDSAIEQGGQGSALAPSPVSGVRREGARGGGRG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MRAHPGGGRCCPEQEEGESAAGGSGAGGDSAIEQGGQGSALAPSPVSGVRREGARGGGRG 60

Qy      61 RGRWKQAGRGGGVCGRGRGRGRGRGRGRGRGRGRGRPPSGGSGLGGDGGGCGGGGSGGGG 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 RGRWKQAGRGGGVCGRGRGRGRGRGRGRGRGRGRGRPPSGGSGLGGDGGGCGGGGSGGGG 120

Qy     121 APRREPVPFPSGSAGPGPRGPRATESGKRMDCPALPPGWKKEEVIRKSGLSAGKSDVYYF 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 APRREPVPFPSGSAGPGPRGPRATESGKRMDCPALPPGWKKEEVIRKSGLSAGKSDVYYF 180

Qy     181 SPSGKKFRSKPQLARYLGNTVDLSSFDFRTGKMMPSKLQKNKQRLRNDPLNQNKGKPDLN 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 SPSGKKFRSKPQLARYLGNTVDLSSFDFRTGKMMPSKLQKNKQRLRNDPLNQNKGKPDLN 240

Qy     241 TTLPIRQTASIFKQPVTKVTNHPSNKVKSDPQRMNEQPRQLFWEKRLQGLSASDVTEQII 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 TTLPIRQTASIFKQPVTKVTNHPSNKVKSDPQRMNEQPRQLFWEKRLQGLSASDVTEQII 300

Qy     301 KTMELPKGLQGVGPGSNDETLLSAVASALHTSSAPITGQVSAAVEKNPAVWLNTSQPLCK 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 KTMELPKGLQGVGPGSNDETLLSAVASALHTSSAPITGQVSAAVEKNPAVWLNTSQPLCK 360

Qy     361 AFIVTDEDIRKQEERVQQVRKKLEEALMADILSRAADTEEMDIEMDSGDEA 411
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 AFIVTDEDIRKQEERVQQVRKKLEEALMADILSRAADTEEMDIEMDSGDEA 411
```

RESULT 2
Q9Z2E1
ID Q9Z2E1 PRELIMINARY; PRT; 414 AA.
AC Q9Z2E1;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Methyl-CpG binding protein MBD2.
GN MBD2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BLACK/6;
RX MEDLINE=98449942; PubMed=9774669;
RA Hendrich B., Bird A.;
RT "Identification and characterization of a family of mammalian methyl-
RT CpG binding proteins.";
RL Mol. Cell. Biol. 18:6538-6547(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=129;
RX MEDLINE=99373255; PubMed=10441743;
RA Hendrich B., Abbott C., McQueen H., Chambers D., Cross S., Bird A.;
RT "Genomic structure and chromosomal mapping of the murine and human
RT mbd1, mbd2, mbd3, and mbd4 genes.";
RL Mamm. Genome 10:906-912(1999).
DR EMBL; AF072243; AAC68872.1; -.
DR EMBL; AF120986; AAD50372.1; -.
DR EMBL; AF120983; AAD50372.1; JOINED.
DR EMBL; AF120984; AAD50372.1; JOINED.
DR EMBL; AF120985; AAD50372.1; JOINED.
DR MGD; MGI:1333813; Mbd2.
DR InterPro; IPR001739; Methyl-CpG_bind.
DR Pfam; PF01429; MBD; 1.
DR SMART; SM00391; MBD; 1.
SQ SEQUENCE 414 AA; 43543 MW; 9601D95E347E8E53 CRC64;

```
  Query Match           96.5%;  Score 2091.5;  DB 11;  Length 414;
  Best Local Similarity  96.6%;  Pred. No. 4.7e-147;
  Matches  400;  Conservative    4;  Mismatches    7;  Indels    3;  Gaps    1;

Qy       1 MRAHPGGGRCCPEQEEGESAAGGSGAGGDSAIEQGGQGSALAPSPVSGVRREGARGGGRG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MRAHPGGGRCCPEQEEGESAAGGSGAGGDSAIEQGGQGSALAPSPVSGVRREGARGGGRG 60

Qy      61 RGRWKQAGRGGGVCGRGRGRGRGRGRGRGRGRGRPPSGGSGLGGDGGGCGGG---GSG 117
           ||||||| |||||||||||||||||||||||| ||||||||||  ||   |||
Db      61 RGRWKQAARGGGVCGRGRGRGRGRGRGRGRGRPQSGGSGLGGDGGGGAGGCGVGSG 120

Qy     118 GGGAPRREPVPFPSGSAGPGPRGPRATESGKRMDCPALPPGWKKEEVIRKSGLSAGKSDV 177
           || ||||:||||||||:||||||||||||||||||||||||||||||||||||||||||||
Db     121 GGVAPRRDPVPFPSGSSGPGPRGPRATESGKRMDCPALPPGWKKEEVIRKSGLSAGKSDV 180

Qy     178 YYFSPSGKKFRSKPQLARYLGNTVDLSSFDFRTGKMMPSKLQKNKQRLRNDPLNQNKGKP 237
           |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db     181 YYFSPSGKKFRSKPQLARYLGNAVDLSSFDFRTGKMMPSKLQKNKQRLRNDPLNQNKGKP 240

Qy     238 DLNTTLPIRQTASIFKQPVTKVTNHPSNKVKSDPQRMNEQPRQLFWEKRLQGLSASDVTE 297
           |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db     241 DLNTTLPIRQTASIFKQPVTKFTNHPSNKVKSDPQRMNEQPRQLFWEKRLQGLSASDVTE 300

Qy     298 QIIKTMELPKGLQGVGPGSNDETLLSAVASALHTSSAPITGQVSAAVEKNPAVWLNTSQP 357
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 QIIKTMELPKGLQGVGPGSNDETLLSAVASALHTSSAPITGQVSAAVEKNPAVWLNTSQP 360

Qy     358 LCKAFIVTDEDIRKQEERVQQVRKKLEEALMADILSRAADTEEMDIEMDSGDEA 411
           |||||||||||||||||||||||||||||||||||||||||:||:|||||||
Db     361 LCKAFIVTDEDIRKQEERVQQVRKKLEEALMADILSRAADTEEVDIDMDSGDEA 414
```

RESULT 3
Q9UIS8
ID Q9UIS8 PRELIMINARY; PRT; 302 AA.
AC Q9UIS8;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Testis-specific methyl-CpG binding protein 2.
GN MBD2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99373255; PubMed=10441743;
RA Hendrich B., Abbott C., McQueen H., Chambers D., Cross S., Bird A.;
RT "Genomic structure and chromosomal mapping of the murine and human
RT Mbd1, Mbd2, Mbd3, and Mbd4 genes.";
RL Mamm. Genome 10:906-912(1999).
DR EMBL; AF120989; AAD56596.1; -.
DR EMBL; AF120988; AAD56596.1; JOINED.
DR InterPro; IPR001739; Methyl-CpG_bind.
DR Pfam; PF01429; MBD; 1.
DR SMART; SM00391; MBD; 1.
SQ SEQUENCE 302 AA; 31745 MW; CCEC65D926222717 CRC64;

```
  Query Match           58.8%;  Score 1274;  DB 4;  Length 302;
  Best Local Similarity  100.0%;  Pred. No. 1.3e-86;
  Matches  234;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MRAHPGGGRCCPEQEEGESAAGGSGAGGDSAIEQGGQGSALAPSPVSGVRREGARGGGRG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MRAHPGGGRCCPEQEEGESAAGGSGAGGDSAIEQGGQGSALAPSPVSGVRREGARGGGRG 60

Qy      61 RGRWKQAGRGGGVCGRGRGRGRGRGRGRGRGRGRPPSGGSGLGGDGGGCGGGGSGGGG 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 RGRWKQAGRGGGVCGRGRGRGRGRGRGRGRGRGRPPSGGSGLGGDGGGCGGGGSGGGG 120

Qy     121 APRREPVPFPSGSAGPGPRGPRATESGKRMDCPALPPGWKKEEVIRKSGLSAGKSDVYYF 180
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 APRREPVPFPSGSAGPGPRGPRATESGKRMDCPALPPGWKKEEVIRKSGLSAGKSDVYYF 180

Qy    181 SPSGKKFRSKPQLARYLGNTVDLSSFDFRTGKMMPSKLQKNKQRLRNDPLNQNK 234
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 SPSGKKFRSKPQLARYLGNTVDLSSFDFRTGKMMPSKLQKNKQRLRNDPLNQNK 234


RESULT 4
Q9Z2D9
ID   Q9Z2D9      PRELIMINARY;      PRT;   249 AA.
AC   Q9Z2D9;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Testis specific methyl-CpG binding protein MBD2.
GN   MBD2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BLACK/6; TISSUE=TESTIS;
RX   MEDLINE=98449942; PubMed=9774669;
RA   Hendrich B., Bird A.;
RT   "Identification and characterization of a family of mammalian methyl-
RT   CpG binding proteins.";
RL   Mol. Cell. Biol. 18:6538-6547(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129;
RX   MEDLINE=99373255; PubMed=10441743;
RA   Hendrich B., Abbott C., McQueen H., Chambers D., Cross S., Bird A.;
RT   "Genomic structure and chromosomal mapping of the murine and human
RT   mbd1, mbd2, mbd3, and mbd4 genes.";
RL   Mamm. Genome 10:906-912(1999).
DR   EMBL; AF072245; AAC68874.1; -.
DR   EMBL; AF120983; AAD50373.1; -.
DR   MGD; MGI:1333813; Mbd2.
DR   InterPro; IPR001739; Methyl-CpG_bind.
DR   Pfam; PF01429; MBD; 1.
DR   SMART; SM00391; MBD; 1.
SQ   SEQUENCE   249 AA;  25493 MW;  530D751E8E5C9540 CRC64;

  Query Match             55.9%;  Score 1210.5;  DB 11;  Length 249;
  Best Local Similarity   95.4%;  Pred. No. 5e-82;
  Matches  226;  Conservative    2;  Mismatches    6;  Indels    3;  Gaps    1;

Qy      1 MRAHPGGGRCCPEQEEGESAAGGSGAGGDSAIEQGGQGSALAPSPVSGVRREGARGGGRG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MRAHPGGGRCCPEQEEGESAAGGSGAGGDSAIEQGGQGSALAPSPVSGVRREGARGGGRG 60

Qy     61 RGRWKQAGRGGGVCGRGRGRGRGRGRGRGRGRGRPPSGGSGLGGDGGGCGGG---GSG 117
          ||||||| |||||||||||||||||||||||||| ||||||||||  ||   |||
Db     61 RGRWKQAARGGGVCGRGRGRGRGRGRGRGRGRGRPQSGGSGLGGDGGGGAGGCGVGSG 120

Qy    118 GGGAPRREPVPFPSGSAGPGPRGPRATESGKRMDCPALPPGWKKEEVIRKSGLSAGKSDV 177
          || ||||:||||||||:||||||||||||||||||||||||||||||||||||||||||||
Db    121 GGVAPRRDPVPFPSGSSGPGPRGPRATESGKRMDCPALPPGWKKEEVIRKSGLSAGKSDV 180

Qy    178 YYFSPSGKKFRSKPQLARYLGNTVDLSSFDFRTGKMMPSKLQKNKQRLRNDPLNQNK 234
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db    181 YYFSPSGKKFRSKPQLARYLGNAVDLSSFDFRTGKMMPSKLQKNKQRLRNDPLNQNK 237


RESULT 5
Q9PUM9
ID   Q9PUM9      PRELIMINARY;      PRT;   282 AA.
AC   Q9PUM9;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Methyl-CpG binding protein MBD3.
GN   MBD3.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Wade P.A., Gégonne A., Jones P.L., Ballestar E., Aubry F.,
RA   Wolffe A.P.;
RT   "The Mi-2 histone deacetylase complex couples DNA methylation to
RT   chromatin remodeling and histone deacetylation.";
RL   Nat. Genet. 0:0-0(1999).
DR   EMBL; AF170346; AAD55389.1; -.
DR   InterPro; IPR001739; Methyl-CpG_bind.
DR   Pfam; PF01429; MBD; 1.
DR   SMART; SM00391; MBD; 1.
SQ   SEQUENCE   282 AA;  31692 MW;  6891EFA05B0371E9 CRC64;

  Query Match             48.9%;  Score 1059;  DB 13;  Length 282;
  Best Local Similarity   75.4%;  Pred. No. 1e-70;
  Matches  202;  Conservative   33;  Mismatches   27;  Indels    6;  Gaps    2;

Qy    148 KRMDCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNTVDLSSFD 207
          || :| ||  ||||||| |:||||||||||||:||||||||||||||||||::|||:||
Db      4 KRWECSAL-QGWKKEEVTRRSGLSAGKSDVYYYSPSGKKFRSKPQLARYLGNSMDLSTFD 62

Qy    208 FRTGKMMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKV 267
          ||||||: ||: ||:||:| | |||:||||||| ||:||||||||||||||||:|||
Db     63 FRTGKMLMSKINKNRQRMRYDGLNQSKGKPDLNTALPVRQTASIFKQPVTKVTNHPTNKV 122

Qy    268 KSDPQRMNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSAVAS 327
          |||||:  :|||||||||:| ||:| |: |:::|||||||||||||||  |||||||:||
Db    123 KSDPQKAVDQPRQLFWEKKLSGLNAFDIAEELVKTMELPKGLQGVGPGCTDETLLSAIAS 182

Qy    328 ALHTSSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEAL 387
          ||||||: |||||:|||||||| |||||||||||||:|||||||||| ||||||||||
Db    183 ALHTSTMPITGQLSAAVEKNPGVWLNTSQPLCKAFMVTDEDIRKQEELVQQVRKKLEEAL 242

Qy    388 MADILSRAADTEE-----MDIEMDSGDE 410
          |||:|:   :  :     :| ::|  :|
Db    243 MADMLAHVEEISKDGGAPLDKDIDDEEE 270


RESULT 6
Q9PUM8
ID   Q9PUM8      PRELIMINARY;      PRT;   303 AA.
AC   Q9PUM8;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Methyl-CpG binding protein MBD3 long form.
GN   MBD3.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Wade P.A., Gegonne A., Jones P.L., Ballestar E., Aubry F.,
RA   Wolffe A.P.;
RT   "The Mi-2 histone deacetylase complex couples DNA methylation to
RT   chromatin remodeling and histone deacetylation.";
RL   Nat. Genet. 0:0-0(1999).
DR   EMBL; AF170347; AAD55390.1; -.
DR   InterPro; IPR001739; Methyl-CpG_bind.
DR   Pfam; PF01429; MBD; 1.
DR   SMART; SM00391; MBD; 1.
SQ   SEQUENCE   303 AA;  34078 MW;  864B727E50EF710B CRC64;

  Query Match             48.8%;  Score 1056.5;  DB 13;  Length 303;
```

```
  Best Local Similarity   70.5%;  Pred. No. 1.7e-70;
  Matches  203;  Conservative   33;  Mismatches   27;  Indels   25;  Gaps    2;

Qy     148 KRMDCPALPPGWKKEEVIRKSGLSAGKSDVYYFS--------------------PSGKKF 187
           || :| ||| ||||||| |:||||||||||||:|                      ||||||
Db       4 KRWECSALPQGWKKEEVTRRSGLSAGKSDVYYYSSSPSRYNRSLRDRVGCLNINPSGKKF 63

Qy     188 RSKPQLARYLGNTVDLSSFDFRTGKMMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQ 247
           |||||||||||::|||:||||||||: ||: ||:||:| | |||:|||||||| ||:||
Db      64 RSKPQLARYLGNSMDLSTFDFRTGKMLMSKINKNRQRMRYDGLNQSKGKPDLNTALPVRQ 123

Qy     248 TASIFKQPVTKVTNHPSNKVKSDPQRMNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPK 307
           |||||||||||||||||:||||||||:  :|||||||||:| ||:| |: |:::|||||||
Db     124 TASIFKQPVTKVTNHPTNKVKSDPQKAVDQPRQLFWEKKLSGLNAFDIAEELVKTMELPK 183

Qy     308 GLQGVGPGSNDETLLSAVASALHTSSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDE 367
           ||||||||  |||||||:|||||||: |||||:|||||||| ||||||||||||||:||||
Db     184 GLQGVGPGCTDETLLSAIASALHTSTMPITGQLSAAVEKNPGVWLNTSQPLCKAFMVTDE 243

Qy     368 DIRKQEERVQQVRKKLEEALMADILSRAADTEE-----MDIEMDSGDE 410
           ||||||| |||||||||||||||:|:   :  :     :| ::|  :|
Db     244 DIRKQEELVQQVRKKLEEALMADMLAHVEEISKDGGAPLDKDIDDEEE 291


RESULT 7
O95983
ID   O95983      PRELIMINARY;   PRT;   291 AA.
AC   O95983;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Methyl-CpG binding protein MBD3.
GN   MBD3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Lamerdin J.E., McCready P.M., Skowronski E., Viswanathan V.,
RA   Burkhart-Schultz K.J., Gordon L., Dias J., Ramirez M., Stilwagen S.,
RA   Phan H., Velasco N., Do L., Regala W., Terry A., Garnes J.,
RA   Danganan L., Erler A., Christensen M., Georgescu A., Avila J., Liu S.,
RA   Attix C., Andreise T., Trankheim M., Amico-Keller G., Coefield J.,
RA   Duarte S., Lucas S., Bruce R., Thomas P., Quan G., Kronmiller B.,
RA   Arellano A., Saunders C., Ow D., Nolan M., Trong S., Kobayashi A.,
RA   Olsen A.S., Carrano A.V.;
RT   "Sequence analysis of a 3.5 Mb contig in human 19p13.3 containing a
RT   serine protease gene cluster.";
RL   Submitted (NOV-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98449942; PubMed=9774669;
RA   Hendrich B., Bird A.;
RT   "Identification and characterization of a family of mammalian methyl-
RT   CpG binding proteins.";
RL   Mol. Cell. Biol. 18:6538-6547(1998).
DR   EMBL; AC005943; AAC72104.1; -.
DR   EMBL; AF072247; AAC68876.1; -.
DR   InterPro; IPR001739; Methyl-CpG_bind.
DR   Pfam; PF01429; MBD; 1.
DR   SMART; SM00391; MBD; 1.
SQ   SEQUENCE   291 AA;  32844 MW;  B62134DD1BEB636B CRC64;

  Query Match           47.5%;  Score 1030;  DB 4;  Length 291;
  Best Local Similarity   72.0%;  Pred. No. 1.5e-68;
  Matches  198;  Conservative   34;  Mismatches   31;  Indels   12;  Gaps    2;

Qy     148 KRMDCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNTVDLSSFD 207
           || :||||| ||::||| |:|||||  ||:|:|||||||||||||||||| ::|||:||
Db       4 KRWECPALPQGWEREEVPRRSGLSAGHRDVFYYSPSGKKFRSKPQLARYLGGSMDLSTFD 63

Qy     208 FRTGKMMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKV 267
           5
```

```
           ||||||: ||: |::||:| |  || |||||||| ||:|||||||||||||:||||||||
Db      64 FRTGKMLMSKMNKSRQRVRYDSSNQVKGKPDLNTALPVRQTASIFKQPVTKITNHPSNKV 123

Qy     268 KSDPQRMNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSAVAS 327
           |||||:  :|||||||||:| ||:| |: |:::|||:|||||||||||  ||||||:||
Db     124 KSDPQKAVDQPRQLFWEKKLSGLNAFDIAEELVKTMDLPKGLQGVGPGCTDETLLSAIAS 183

Qy     328 ALHTSSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEAL 387
           |||||: |||||:|||||||| |||||:|||||||:||||||||||  ||||||:|||||
Db     184 ALHTSTMPITGQLSAAVEKNPGVWLNTTQPLCKAFMVTDEDIRKQEELVQQVRKRLEEAL 243

Qy     388 MADILSR----AADTE--------EMDIEMDSGDE 410
           |||:|:     | | |         | | | |  :|
Db     244 MADMLAHVEELARDGEAPLDKACAEDDDEEDEEEE 278


RESULT 8
Q9Z2D8
ID   Q9Z2D8      PRELIMINARY;      PRT;   285 AA.
AC   Q9Z2D8;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Methyl-CpG binding  protein MBD3.
GN   MBD3.
OS   Mus musculus (Mouse), and
OS   Mus musculus domesticus (western European house mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090, 10092;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98449942; PubMed=9774669;
RA   Hendrich B., Bird A.;
RT   "Identification and characterization of a family of mammalian methyl-
RT   CpG binding proteins.";
RL   Mol. Cell. Biol. 18:6538-6547(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129;
RX   MEDLINE=99373255; PubMed=10441743;
RA   Hendrich B., Abbott C., McQueen H., Chambers D., Cross S., Bird A.;
RT   "Genomic structure and chromosomal mapping of the murine and human
RT   mbd1, mbd2, mbd3, and mbd4 genes.";
RL   Mamm. Genome 10:906-912(1999).
DR   EMBL; AF072248; AAC68877.1; -.
DR   EMBL; AF120995; AAD48909.1; -.
DR   MGD; MGI:1333812; Mbd3.
DR   InterPro; IPR001739; Methyl-CpG_bind.
DR   Pfam; PF01429; MBD; 1.
DR   SMART; SM00391; MBD; 1.
SQ   SEQUENCE   285 AA;  32168 MW;  E4E57BD48463643F CRC64;

  Query Match           47.1%;  Score 1021.5;  DB 11;  Length 285;
  Best Local Similarity   75.2%;  Pred. No. 6.1e-68;
  Matches  191;  Conservative   35;  Mismatches   25;  Indels    3;  Gaps    1;

Qy     148 KRMDCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNTVDLSSFD 207
           || :||||| ||::||| |:|||||  ||:|:|||||||||||||||||| ::|||:||
Db       4 KRWECPALPQGWEREEVPRRSGLSAGHRDVFYYSPSGKKFRSKPQLARYLGGSMDLSTFD 63

Qy     208 FRTGKMMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKV 267
           |||||: :|: |::||:| |  || |||||||| ||:|||||||||||||:||||||||
Db      64 FRTGKMLMNKMNKSRQRVRYDSSNQVKGKPDLNTALPVRQTASIFKQPVTKITNHPSNKV 123

Qy     268 KSDPQRMNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSAVAS 327
           |||||:  :|||||||||:| |||| |: |:::|||:|||||||||||  ||||||:||
Db     124 KSDPQKAVDQPRQLFWEKKLSGLSAFDIAEELVRTMDLPKGLQGVGPGCTDETLLSAIAS 183

Qy     328 ALHTSSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEAL 387
           |||||: |||||:|||||||| |||||:|||||||:|||:|||||||  ||||||:|||||
Db     184 ALHTSTLPITGQLSAAVEKNPGVWLNTAQPLCKAFMVTDDDIRKQEELVQQVRKRLEEAL 243
```

```
Qy    388 MADILSRAADTEEM 401
          |||:|   |  ||:
Db    244 MADML---AHVEEL 254


RESULT 9
O60535
ID   O60535      PRELIMINARY;      PRT;   200 AA.
AC   O60535;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Antigen NY-CO-41 (Fragment).
GN   NY-CO-41.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=COLON CANCER METASTASIS TO LUNG;
RX   MEDLINE=98272252; PubMed=9610721;
RA   Scanlan M.J., Chen Y.T., Williamson B., Gure A.O., Stockert E.,
RA   Gordan J.D., Tureci O., Sahin U., Pfreundschuh M., Old L.J.;
RT   "Characterization of human colon cancer antigens recognized by
RT   autologous antibodies.";
RL   Int. J. Cancer 76:652-658(1998).
DR   EMBL; AF039701; AAC18050.1; -.
FT   NON_TER      1      1
SQ   SEQUENCE   200 AA;  22270 MW;  A70B46B35CA55AFD CRC64;

  Query Match             46.7%;  Score 1013;  DB 4;  Length 200;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-67;
  Matches  200;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    212 KMMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKVKSDP 271
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 KMMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKVKSDP 60

Qy    272 QRMNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSAVASALHT 331
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 QRMNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSAVASALHT 120

Qy    332 SSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEALMADI 391
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 SSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEALMADI 180

Qy    392 LSRAADTEEMDIEMDSGDEA 411
          ||||||||||||||||||||
Db    181 LSRAADTEEMDIEMDSGDEA 200


RESULT 10
Q9VHB7
ID   Q9VHB7      PRELIMINARY;      PRT;   314 AA.
AC   Q9VHB7;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Methyl-CpG-binding-domain-like-protein.
GN   METHYL-CPG-BINDING-DOMAIN-LIKE-PROTEIN OR CG8208.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003683; AAF54400.1; -.
DR   FlyBase; FBgn0027950; methyl-CpG-binding-domain-like-protein.
DR   InterPro; IPR001739; Methyl-CpG_bind.
DR   Pfam; PF01429; MBD; 1.
DR   SMART; SM00391; MBD; 1.
SQ   SEQUENCE   314 AA;  33803 MW;  7B6BDFE63873D230 CRC64;

  Query Match             20.9%;  Score 452.5;  DB 5;  Length 314;
  Best Local Similarity   35.2%;  Pred. No. 1e-25;
  Matches  112;  Conservative   48;  Mismatches   77;  Indels   81;  Gaps    9;

Qy    141 PRATESGKRMDCPALPPGWKKEEVIRKSGLSA-------------------GKSDVYYF 180
          |  |   ||:||  || ||:::|| |||| ||                    | ||:|:
Db      5 PSVTIERKRVDCSVLPKGWQRDEV-RKSGSSANNNASSNNNSSATASSNNNNNKVDVFYY 63

Qy    181 SPSGKKFRSKPQLARYLGNTVDLSSFDFRTGKM-----------------MP-------- 215
          ||:||:   |||         |::  ||: |||                 ||
Db     64 SPTGKRAEGKPQ---------DIAIPDFQPGKMPHCALPSPSISLYRCSAMPLPIASGGG 114

Qy    216 ------SKLQKNKQRLRNDPLNQNKG--------------KPDLNTTLPIRQTASIFKQ 254
                |     |::          |               : |::   ||||||||||
Db    115 NGATSGSAANALKRKFARSQGGNAAGAAGAAPPAATASRALRTDVSLVPPIRQTASIFKQ 174

Qy    255 PVTKVTNHPSN--KVKSDPQR-MNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQG 311
          ||| : ||  :  | |::|:    |:|:|||||||: | |   : : :  : ||| ::
Db    175 PVTVIRNHKQDPAKAKNEPKHGTREKPKQLFWEKRLERLRACHDSGEELDDISLPKTIRT 234

Qy    312 VGPGSNDETLLSAVASALHTSSAPITGQVS--AAVEKNPAVWLNTSQPLCKAFIVTDEDI 369
          |||  |::|:| :||:|||  :| : || |  | : ||   ::|  |||  | |::::||
Db    235 VGPNVNEQTVLQSVATALHMLNAGVHGQSSTKADLTKNAMAFMNPEQPLMHAVIISEDDI 294

Qy    370 RKQEERVQQVRKKLEEAL 387
          ||||:||   |:||::||
Db    295 RKQEDRVGVARRKLQDAL 312


RESULT 11
O95242
```

```
ID   O95242      PRELIMINARY;      PRT;   146 AA.
AC   O95242;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Testis specific methyl-CpG binding protein MBD2 (Fragment).
GN   MBD2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=TESTIS;
RX   MEDLINE=98449942; PubMed=9774669;
RA   Hendrich B., Bird A.;
RT   "Identification and characterization of a family of mammalian methyl-
RT   CpG binding proteins.";
RL   Mol. Cell. Biol. 18:6538-6547(1998).
DR   EMBL; AF072246; AAC68875.1; -.
DR   InterPro; IPR001739; Methyl-CpG_bind.
DR   Pfam; PF01429; MBD; 1.
DR   SMART; SM00391; MBD; 1.
FT   NON_TER       1      1
SQ   SEQUENCE   146 AA;  16931 MW;  9D6CC3CFE140EEEA CRC64;

  Query Match             18.9%;  Score 409;  DB 4;  Length 146;
  Best Local Similarity   100.0%;  Pred. No. 6.5e-23;
  Matches   78;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    157 PGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNTVDLSSFDFRTGKMMPS 216
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 PGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNTVDLSSFDFRTGKMMPS 60

Qy    217 KLQKNKQRLRNDPLNQNK 234
          ||||||||||||||||||
Db     61 KLQKNKQRLRNDPLNQNK 78


RESULT 12
Q9V424
ID   Q9V424      PRELIMINARY;      PRT;   226 AA.
AC   Q9V424;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Methyl-CpG-binding-domain-like-protein.
GN   METHYL-CPG-BINDING-DOMAIN-LIKE-PROTEIN OR CG8208.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Wade P.A., Gegonne A., Jones P.L., Ballestar E., Aubry F.,
RA   Wolffe A.P.;
RT   "The Mi-2 histone deacetylase complex couples DNA methylation to
RT   chromatin remodeling and histone deacetylation.";
RL   Nat. Genet. 0:0-0(1999).
DR   EMBL; AE003683; AAF54401.1; -.
DR   EMBL; AF171098; AAD55391.1; -.
DR   FlyBase; FBgn0027950; methyl-CpG-binding-domain-like-protein.
DR   InterPro; IPR001739; Methyl-CpG_bind.
DR   Pfam; PF01429; MBD; 1.
DR   SMART; SM00391; MBD; 1.
SQ   SEQUENCE   226 AA;  25292 MW;  26F2DD296442B4D3 CRC64;

  Query Match             18.5%;  Score 400.5;  DB 5;  Length 226;
  Best Local Similarity   38.5%;  Pred. No. 4.9e-22;
  Matches   97;  Conservative   43;  Mismatches   75;  Indels   37;  Gaps    7;

Qy    141 PRATESGKRMDCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNT 200
          |  |   ||:||  || ||:::|| |||| ||       : |     |    :   |
Db      5 PSVTIERKRVDCSVLPKGWQRDEV-RKSGSSANN------NASSNNNSSATASSNNNNNK 57

Qy    201 VDLSSFDFRTGKMMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTASIFKQPVTKVT 260
          ||:  :              : ||          |::   |||||||||||||| :
Db     58 VDVFYY--------------SRALRT----------DVSLVPPIRQTASIFKQPVTVIR 92

Qy    261 NHPSN--KVKSDPQR-MNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSN 317
          ||  :  | |::|:    |:|:|||||||: | |   : : :  : ||| :: |||  |
Db     93 NHKQDPAKAKNEPKHGTREKPKQLFWEKRLERLRACHDSGEELDDISLPKTIRTVGPNVN 152

Qy    318 DETLLSAVASALHTSSAPITGQVS--AAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEER 375
          ::|:| :||:|||  :| : || |  | : ||   ::|  |||  | |::::|||||:|
Db    153 EQTVLQSVATALHMLNAGVHGQSSTKADLTKNAMAFMNPEQPLMHAVIISEDDIRKQEDR 212

Qy    376 VQQVRKKLEEAL 387
          |   |:||::||
Db    213 VGVARRKLQDAL 224


RESULT 13
Q8WWY6
ID   Q8WWY6      PRELIMINARY;      PRT;   194 AA.
AC   Q8WWY6;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Methyl-CpG binding domain protein 3-like.
GN   MBD3L.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Jiang C.L., Szabo P.E., O'Connor T.R., Mann J.R., Pfeifer G.P.;
RT   "MBD3L, a gene homologous to the methyl-CpG-binding domain protein
RT   genes MBD3 and MBD2, is expressed specifically in postmeiotic cells of
RT   the testis.";
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY038022; AAK72591.1; -.
SQ   SEQUENCE   194 AA;  21562 MW;  F9EC132569145DC9 CRC64;

  Query Match             15.6%;  Score 339;  DB 4;  Length 194;
  Best Local Similarity   41.8%;  Pred. No. 1.5e-17;
  Matches   76;  Conservative   39;  Mismatches   65;  Indels    2;  Gaps    2;

Qy        218 LQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKVKSDP-QRMNE 276
              : |: || : | :|| | || |:|::|:| ::  ||:|||::| || |:|:    :   |
Db          1 MAKSSQRKQRDCVNQCKSKPGLSTSIPLRMSSYTFKRPVTRITPHPGNEVRYHQWEESLE 60

Qy        277 QPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSAVASAL-HTSSAP 335
              :|:|: |::||||| |     ::   |::|    || : |      :||  :|| | |:   |
Db         61 KPQQVCWQRRLQGLQAYSSAGELSSTLDLANTLQKLVPSYTGGSLLEDLASGLEHSCPMP 120

Qy        336 ITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEALMADILSRA 395
                 | |||  ||  :  || ||| |:||:|||||| :|: ||::|  ||:|| |:
Db        121 HLACSSDAVEIIPAEGVGISQLLCKQFLVTEEDIRKQEGKVKTVRERLAIALIADGLANE 180

Qy        396 AD 397
              |:
Db        181 AE 182


RESULT 14
Q9IPQ9
ID   Q9IPQ9      PRELIMINARY;      PRT;   619 AA.
AC   Q9IPQ9;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   EBNA-1.
GN   EBNA-1.
OS   Cynomolgus Epstein-Barr Virus TsB-B6.
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Gammaherpesvirinae; Lymphocryptovirus.
OX   NCBI_TaxID=104240;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20428646; PubMed=10971128;
RA   Ohara N., Hayashi K., Teramoto N., Oka T., Fujimoto K., Yoshikawa Y.,
RA   Castanos-Velez E., Biberfeld P., Akagi T.;
RT   "Sequence Analysis and Variation of EBNA-1 in Epstein-Barr Virus-
RT   Related Herpesvirus of Cynomolgus Monkey.";
RL   Intervirology 43:102-106(2000).
DR   EMBL; AB037275; BAB03281.1; -.
DR   HSSP; P03211; 1VHI.
DR   InterPro; IPR004186; EBNA1.
DR   InterPro; IPR002952; Eggshell.
DR   Pfam; PF02905; EBNA1; 1.
DR   PRINTS; PR01561; EDG8RECEPTOR.
DR   PRINTS; PR01228; EGGSHELL.
DR   PRINTS; PR01574; TUBBYPROTEIN.
SQ   SEQUENCE   619 AA;  56431 MW;  CF02010FA3D92972 CRC64;

  Query Match             12.5%;  Score 271;  DB 12;  Length 619;
  Best Local Similarity   37.6%;  Pred. No. 7.2e-12;
  Matches   85;  Conservative    8;  Mismatches   69;  Indels   64;  Gaps   10;

Qy         17 GESAAGGSGAGGDSAIEQGGQGSALAPSPVSGVRREGARGGGRGRGRWKQAGRG-----G 71
              |   ||||||||  |   || |: |    ||     |  ||||||||    |||     |
Db        236 GAGGAGGSGAGGSGAGGAGGSGAGGAGG--SGAGGAGGSGGGRGRGRGGSRGRGGSRGRG 293

Qy         72 GVCGRGRGRG---RGRGRGRGRGRGRG----------RPPSGGSGLG-----GDGGGCGG 113
```

```
                |  |||||||   |||||||||||||                 | ||| |  |      |
Db        294 GSRGRGRGRGGGSRGRGRGRGRGRGRGEGPRQGEKRPRSPSGSSSSGSSSRSSSSGRASS 353

Qy        114 GGSGGGGA-------------PRREPVPFPSGSAGPGPRGPRATESGKRMDCPALPPGWK 160
              |||  |||             |   ||   |||    ||    |: |        |||
Db        354 GGSSSGGALSNGSFYGFPGDRPLTVTVP---GSALGRYRGTDGTDGGDE------PPGAM 404

Qy        161 KE------------EVIRKSGLSAGKSDVYYFSPSG-----KKFRS 189
              ::            :     |   ||   ::    |      |||:|
Db        405 EQGPEEDPGEGPSRQHTTSGGRGGGKKGGWFGRHRGEGRGNKKFQS 450


RESULT 15
Q9IPQ8
ID   Q9IPQ8      PRELIMINARY;      PRT;   588 AA.
AC   Q9IPQ8;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   EBNA-1.
GN   EBNA-1.
OS   Cynomolgus Epstein-Barr Virus Si-IIA.
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Gammaherpesvirinae; Lymphocryptovirus.
OX   NCBI_TaxID=114769;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20428646; PubMed=10971128;
RA   Ohara N., Hayashi K., Teramoto N., Oka T., Fujimoto K., Yoshikawa Y.,
RA   Castanos-Velez E., Biberfeld P., Akagi T.;
RT   "Sequence Analysis and Variation of EBNA-1 in Epstein-Barr Virus-
RT   Related Herpesvirus of Cynomolgus Monkey.";
RL   Intervirology 43:102-106(2000).
DR   EMBL; AB037276; BAB03282.1; -.
DR   HSSP; P03211; 1VHI.
DR   InterPro; IPR004186; EBNA1.
DR   InterPro; IPR002952; Eggshell.
DR   Pfam; PF02905; EBNA1; 1.
DR   PRINTS; PR01561; EDG8RECEPTOR.
DR   PRINTS; PR01228; EGGSHELL.
SQ   SEQUENCE   588 AA;  54634 MW;  CC532F2AE8B816EE CRC64;

  Query Match             12.2%;  Score 264.5;  DB 12;  Length 588;
  Best Local Similarity   36.7%;  Pred. No. 2e-11;
  Matches   83;  Conservative    7;  Mismatches   77;  Indels   59;  Gaps   10;

Qy         17 GESAAGGSGAGGDSAIEQGGQGSALAPSPVSGVRREGARGGGRGRGRWKQAGRG------ 70
              |   ||||||||  |:  || |:  |    ||     |  ||| ||||    |||
Db        200 GAVGAGGSGAGGSGAVGAGGSGAGGAGG--SGAGGAGGSGGGGRGRGGSRGRGGSRGRG 257

Qy         71 -----GGVCGRGRGRG--RGRGRGRGRGRG----------RPPSGGSGLG------GD 107
                   ||  |||||||  |||||||||||||           | ||| |  |
Db        258 GSRGRGGSRGRGRGRGGSRGRGRGRGRGRGQGPRQGEKRPRSPSGSSSSGSSSRSSSS 317

Qy        108 GGGCGGGGSGGG--------GAPRREP--VPFPSGSAG---------PGPRGPRATESGK 148
              |    || | ||          ||   |  |  |  | : |          |   | | | |
Db        318 GRASSGGSSSGGTLSNGSFYGFPGDRPLTVTVPGSALGRYRGTDGTDGGDEPPGAMEQGP 377

Qy        149 RMDCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSG-----KKFRS 189
              |   |       :        |  ||    ::   |      |||:|
Db        378 EED----PGEGPSRQHTTSGGRGGGKKGGWFGRHRGEGRGNKKFQS 419


Search completed: March 12, 2003, 09:12:40
Job time : 66.7949 secs
```

GenCore version 5.1.4_p5_4578

OM protein - protein search, using sw model

Run on: March 12, 2003, 05:40:46 ; Search time 19.5423 Seconds
(without alignments)
618.801 Million cell updates/sec

Title: US-09-554-414B-2
Perfect score: 2167
Sequence: 1 MRAHPGGGRCCPEQEEGESA.....LSRAADTEEMDIEMDSGDEA 411

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1344	62.0	263	4	US-09-149-476-580	Sequence 580, App
2	214.5	9.9	574	4	US-09-079-431B-6	Sequence 6, Appli
3	214.5	9.9	629	4	US-09-079-431B-4	Sequence 4, Appli
4	214.5	9.9	630	4	US-09-079-431B-2	Sequence 2, Appli
5	214	9.9	641	4	US-09-249-585A-3	Sequence 3, Appli
6	210	9.7	235	2	US-08-529-190B-1	Sequence 1, Appli
7	209.5	9.7	201	4	US-09-052-995-1	Sequence 1, Appli
8	209.5	9.7	201	4	US-09-053-003-40	Sequence 40, Appl
9	209.5	9.7	201	4	US-09-054-281-22	Sequence 22, Appl
10	209.5	9.7	201	4	US-09-478-948-6	Sequence 6, Appli
11	196.5	9.1	1136	4	US-08-806-029-9	Sequence 9, Appli
12	196.5	9.1	1177	1	US-07-609-716-31	Sequence 31, Appl
13	196.5	9.1	1177	1	US-08-175-155-29	Sequence 29, Appl
14	196.5	9.1	1177	1	US-08-477-509B-64	Sequence 64, Appl
15	196.5	9.1	1177	2	US-08-707-237A-35	Sequence 35, Appl
16	196.5	9.1	1177	3	US-08-482-085B-64	Sequence 64, Appl
17	196.5	9.1	1177	4	US-08-475-411A-31	Sequence 31, Appl
18	196.5	9.1	1177	4	US-08-478-029A-31	Sequence 31, Appl
19	196.5	9.1	1177	4	US-09-444-791A-64	Sequence 64, Appl
20	195.5	9.0	1332	1	US-07-609-716-41	Sequence 41, Appl
21	195.5	9.0	1332	4	US-08-475-411A-41	Sequence 41, Appl
22	195.5	9.0	1332	4	US-08-478-029A-41	Sequence 41, Appl
23	195	9.0	745	2	US-09-010-928B-28	Sequence 28, Appl
24	195	9.0	870	2	US-09-010-928B-2	Sequence 2, Appli
25	193	8.9	784	1	US-07-609-716-48	Sequence 48, Appl
26	193	8.9	784	4	US-08-475-411A-48	Sequence 48, Appl
27	193	8.9	784	4	US-08-478-029A-48	Sequence 48, Appl
28	192.5	8.9	649	1	US-07-609-716-49	Sequence 49, Appl
29	192.5	8.9	649	4	US-08-475-411A-49	Sequence 49, Appl
30	192.5	8.9	649	4	US-08-478-029A-49	Sequence 49, Appl
31	192.5	8.9	718	1	US-08-425-069-2	Sequence 2, Appli
32	192.5	8.9	718	2	US-08-317-844B-2	Sequence 2, Appli
33	192.5	8.9	747	3	US-09-034-177-3	Sequence 3, Appli
34	192.5	8.9	945	1	US-08-089-862-6	Sequence 6, Appli
35	192.5	8.9	945	1	US-08-587-333-13	Sequence 13, Appl
36	192.5	8.9	945	5	PCT-US94-07776-11	Sequence 11, Appl
37	192.5	8.9	1059	1	US-08-175-155-48	Sequence 48, Appl
38	192.5	8.9	1059	2	US-08-707-237A-54	Sequence 54, Appl
39	192.5	8.9	1059	4	US-08-806-029-10	Sequence 10, Appl
40	192.5	8.9	1101	1	US-08-477-509B-83	Sequence 83, Appl
41	192.5	8.9	1101	3	US-08-482-085B-83	Sequence 83, Appl
42	192.5	8.9	1101	4	US-09-444-791A-83	Sequence 83, Appl
43	189.5	8.7	651	4	US-08-556-978B-19	Sequence 19, Appl
44	189.5	8.7	651	4	US-09-247-806-1	Sequence 1, Appli
45	189	8.7	1958	1	US-07-945-283-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-09-149-476-580
; Sequence 580, Application US/09149476
; Patent No. 6420526
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 186 Human Secreted proteins
; FILE REFERENCE: PZ002P1
; CURRENT APPLICATION NUMBER: US/09/149,476
; CURRENT FILING DATE: 1998-09-08
; EARLIER APPLICATION NUMBER: PCT/US98/04493
; EARLIER FILING DATE: 1998-03-06
; EARLIER APPLICATION NUMBER: 60/040,162
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,333
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/038,621
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,626
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,334
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,336
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,163
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/047,600
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,615
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,597
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,502
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,633
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,583
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,617
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,618
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,503
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,592
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,581
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,584
; EARLIER FILING DATE: 1997-05-23

; EARLIER APPLICATION NUMBER: 60/047,500
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,587
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,492
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,598
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,613
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,582
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,596
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,612
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,632
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,601
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/043,580
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,568
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,314
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,569
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,311
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,671
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,674
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,669
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,312
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,313
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,672
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,315
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/048,974
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/056,886
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,877
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,889
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,893
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,630
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,878
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,662
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,872
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,882
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,637
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,903
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,888
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,879
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,880
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,894
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,911
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,636
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,874
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,910
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,864
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,631
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,845
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,892
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/057,761
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/047,595
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,599
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,588
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,585
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,586
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,590
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,594
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,589
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,593
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,614
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/043,578
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,576
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/047,501
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/043,670
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/056,632
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,664
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,876
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,881
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,909
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,875
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,862
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,887
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,908
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/048,964
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/057,650
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: 60/056,884
; EARLIER FILING DATE: 1997-08-22

```
;  EARLIER APPLICATION NUMBER: 60/057,669
;  EARLIER FILING DATE: 1997-09-05
;  EARLIER APPLICATION NUMBER: 60/049,610
;  EARLIER FILING DATE: 1997-06-13
;  EARLIER APPLICATION NUMBER: 60/061,060
;  EARLIER FILING DATE: 1997-10-02

  Query Match             62.0%;  Score 1344;  DB 4;  Length 263;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-98;
  Matches  262;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    150 MDCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNTVDLSSFDFR 209
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MDCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNTVDLSSFDFR 60

Qy    210 TGKMMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKVKS 269
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TGKMMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKVKS 120

Qy    270 DPQRMNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSAVASAL 329
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 DPQRMNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSAVASAL 180

Qy    330 HTSSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEALMA 389
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 HTSSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEALMA 240

Qy    390 DILSRAADTEEMDIEMDSGDEA 411
          ||||||||||||||||||||||
Db    241 DILSRAADTEEMDIEMDSGDEA 262


RESULT 2
US-09-079-431B-6
; Sequence 6, Application US/09079431B
; Patent No. 6326484
; GENERAL INFORMATION:
;  APPLICANT: Gage, Fred
;  APPLICANT:  Ueba, Tetsuya
;  TITLE OF INVENTION: TRANSCRIPTION FACTOR REGULATING FGF-2
;  TITLE OF INVENTION:  AND VARIANTS THEREOF
;  FILE REFERENCE: SALKINS.015A
;  CURRENT APPLICATION NUMBER: US/09/079,431B
;  CURRENT FILING DATE:  1998-05-14
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6
;   LENGTH: 574
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-079-431B-6

  Query Match             9.9%;  Score 214.5;  DB 4;  Length 574;
  Best Local Similarity   45.7%;  Pred. No. 4.6e-09;
  Matches   42;  Conservative   13;  Mismatches   30;  Indels    7;  Gaps    1;

Qy    150 MDCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNTVDLSSFDFR 209
          :|||||  ||||: || |||| : |:|| || ||:| : ||| :| ||||   ||: |||:
Db      6 LDCPALGPGWKRREVFRKSGATCGRSDTYYQSPTGDRIRSKVELTRYLGPACDLTLFDFK 65

Qy    210 TG-------KMMPSKLQKNKQRLRNDPLNQNK 234
          ||       | |  :   |::  : |    |
Db     66 QGILCYPAPKAHPVAVASKKRKKPSRPAKTRK 97


RESULT 3
US-09-079-431B-4
; Sequence 4, Application US/09079431B
; Patent No. 6326484
; GENERAL INFORMATION:
;  APPLICANT: Gage, Fred
;  APPLICANT:  Ueba, Tetsuya
;  TITLE OF INVENTION: TRANSCRIPTION FACTOR REGULATING FGF-2
;  TITLE OF INVENTION:  AND VARIANTS THEREOF
;  FILE REFERENCE: SALKINS.015A
;  CURRENT APPLICATION NUMBER: US/09/079,431B
;  CURRENT FILING DATE:  1998-05-14
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
;   LENGTH: 629
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-079-431B-4

  Query Match             9.9%;  Score 214.5;  DB 4;  Length 629;
  Best Local Similarity   45.7%;  Pred. No. 5.2e-09;
  Matches   42;  Conservative   13;  Mismatches   30;  Indels    7;  Gaps    1;

Qy    150 MDCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNTVDLSSFDFR 209
          :|||||  ||||: || |||| : |:|| || ||:| : ||| :| ||||   ||: |||:
Db      6 LDCPALGPGWKRREVFRKSGATCGRSDTYYQSPTGDRIRSKVELTRYLGPACDLTLFDFK 65

Qy    210 TG-------KMMPSKLQKNKQRLRNDPLNQNK 234
          ||       | |  :   |::  : |    |
Db     66 QGILCYPAPKAHPVAVASKKRKKPSRPAKTRK 97


RESULT 4
US-09-079-431B-2
; Sequence 2, Application US/09079431B
; Patent No. 6326484
; GENERAL INFORMATION:
;  APPLICANT: Gage, Fred
;  APPLICANT:  Ueba, Tetsuya
;  TITLE OF INVENTION: TRANSCRIPTION FACTOR REGULATING FGF-2
;  TITLE OF INVENTION:  AND VARIANTS THEREOF
;  FILE REFERENCE: SALKINS.015A
;  CURRENT APPLICATION NUMBER: US/09/079,431B
;  CURRENT FILING DATE:  1998-05-14
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 630
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-079-431B-2

  Query Match             9.9%;  Score 214.5;  DB 4;  Length 630;
  Best Local Similarity   45.7%;  Pred. No. 5.2e-09;
  Matches   42;  Conservative   13;  Mismatches   30;  Indels    7;  Gaps    1;

Qy    150 MDCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNTVDLSSFDFR 209
          :|||||  ||||: || |||| : |:|| || ||:| : ||| :| ||||   ||: |||:
Db      6 LDCPALGPGWKRREVFRKSGATCGRSDTYYQSPTGDRIRSKVELTRYLGPACDLTLFDFK 65

Qy    210 TG-------KMMPSKLQKNKQRLRNDPLNQNK 234
          ||       | |  :   |::  : |    |
Db     66 QGILCYPAPKAHPVAVASKKRKKPSRPAKTRK 97


RESULT 5
US-09-249-585A-3
; Sequence 3, Application US/09249585A
; Patent No. 6417002
; GENERAL INFORMATION:
;  APPLICANT: Horlick, Robert
;  TITLE OF INVENTION: METHOD FOR MAINTENANCE AND SELECTION OF EPISOMES
;  FILE REFERENCE: 0867/0D905
;  CURRENT APPLICATION NUMBER: US/09/249,585A
;  CURRENT FILING DATE:  1999-02-11
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
;   LENGTH: 641
;   TYPE: PRT
```

```
;   ORGANISM: Epstein Barr Virus
US-09-249-585A-3

  Query Match             9.9%;  Score 214;  DB 4;  Length 641;
  Best Local Similarity   39.7%;  Pred. No. 5.8e-09;
  Matches   75;  Conservative    1;  Mismatches   57;  Indels   56;  Gaps   11;

Qy       3 AHPGGGRCCPEQEEGESAAGGSGAGGDSAIEQGGQGSALAPSPVSGVRREGARGGGRGRG 62
           |  |||         |   |||:||||      || | | |     |    ||  || |||
Db     278 AGAGGG----AGGAGAGGAGGAGAGGAGGAGAGGAGGAGAG---GGAGAGGAGAGGGGRG 330

Qy      63 RWKQAGRG-GGVCGRGRG-----RGRG----------RGRGRGRGRGRGRPPSGGSGLGG 106
           |    ||| ||  |||||     ||||          | ||||||||  || |  |
Db     331 RGGSGGRGRGGSGGRGRGGSGGRRGRGRERARGGSRERARGRGRGRGEKRPRSPSS---- 386

Qy     107 DGGGCGGGGSGGGGAPRREPV---PF--PSGSAG-------PGPRG-----PRATESGKR 149
                   |  |  ||| |    ||  | | |         || |     | | | |
Db     387 -------QSSSSGSPPRRPPPGRRPFFHPVGEADYFEYHQEGGPDGEPDVPPGAIEQG-- 437

Qy     150 MDCPALPPG 158
             ||  ||
Db     438 ---PADDPG 443


RESULT 6
US-08-529-190B-1
; Sequence 1, Application US/08529190B
; Patent No. 5833991
;  GENERAL INFORMATION:
;    APPLICANT:  Masucci, Maria G.
;    TITLE OF INVENTION:  GLYCINE-CONTAINING SEQUENCES
;    TITLE OF INVENTION:  CONFERRING INVISIBILITY TO THE IMMUNE SYSTEM
;    NUMBER OF SEQUENCES:  76
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Banner & Witcoff, Ltd.
;      STREET:  One Financial Center
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  Wordperfect 6.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/529,190B
;      FILING DATE:  15-SEP-1995
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  SE9501324-9
;      FILING DATE:  10-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US08/522,595
;      FILING DATE:  01-SEP-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Williams, Ph.D., Kathleen A
;      REGISTRATION NUMBER:  34,380
;      REFERENCE/DOCKET NUMBER:  3255/53015
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-345-9100
;      TELEFAX:  617-345-9111
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  235 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-529-190B-1

  Query Match             9.7%;  Score 210;  DB 2;  Length 235;
  Best Local Similarity   43.5%;  Pred. No. 3.2e-09;
  Matches   54;  Conservative    4;  Mismatches   54;  Indels   12;  Gaps    3;

Qy      17 GESAAGGSGAGGDSAIEQGGQGSALAPSPVSGVRREGARGGGRGRGRWKQAGRGGGVCGR 76
           |   |||:||||  |   || | | |      |    || ||    |    || |||  |
Db      46 GAGGAGGAGAGG-GAGAGGGAGGAGAGGGAGGAGGAGAGGGAGAGGAGGAGAGGGAGGA 104

Qy      77 GRGRGRGRGRGRGRGRGRGRPPSGGSGLGGDGGGCGGGGSGGGGAPRREPVPFPSGSAGP 136
           | | | | | | |   | |   :||:| ||  || || |:|| ||          | ||
Db     105 G-GAGAGGGAGAGGAGGAGAGGAGAGGGAGGAGGAGAGGAGA----------GGAGA 153

Qy     137 GPRG 140
           | |
Db     154 GGAG 157


RESULT 7
US-09-052-995-1
; Sequence 1, Application US/09052995
; Patent No. 6183956
;  GENERAL INFORMATION:
;    APPLICANT:  Sivaraja, Mohanram
;    APPLICANT:  Strulovici, Berta
;    APPLICANT:  Flores, Osvaldo A.
;    TITLE OF INVENTION:  High Throughput In Vitro Screening Assay
;    TITLE OF INVENTION:  for Transcription Modulators
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend and Crew LLP
;      STREET:  Two Embarcadero Center, Eighth Floor
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-3834
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/052,995
;      FILING DATE:  31-MAR-1998
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parent, Annette S.
;      REGISTRATION NUMBER:  42,058
;      REFERENCE/DOCKET NUMBER:  018781-000600US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 576-0200
;      TELEFAX:  (415) 576-0300
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  201 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-052-995-1

  Query Match             9.7%;  Score 209.5;  DB 4;  Length 201;
  Best Local Similarity   40.7%;  Pred. No. 2.9e-09;
  Matches   55;  Conservative    0;  Mismatches   63;  Indels   17;  Gaps    2;

Qy       6 GGGRCCPEQEEGESAAGGSGAGGDSAIEQGGQGSALAPSPVSGVRREGARGGGRGRGRWK 65
           |||         |    || | ||      || |         |    |  ||| | |
Db      17 GGG--------GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 68

Qy      66 QAGRGGGVCGRGRGRGRGRGRGRGRGRGRPPSGGSGLGGDGGGCGGGGSGGGGAPRRE 125
             | |||  | | | | | | | | | | | | | | | | || | || ||| |||| ||||
Db      69 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGPGGGGGGGGGGGGGGGGGGGGGG---- 124

Qy     126 PVPFPSGSAGPGPRG 140
               | | | |
Db     125 -----GGGGGGGGG 134
```

```
RESULT 8
US-09-053-003-40
; Sequence 40, Application US/09053003
; Patent No. 6207391
;  GENERAL INFORMATION:
;    APPLICANT:  Wu, Pengguang
;    APPLICANT:  McKinney, Judi
;    TITLE OF INVENTION:  High-Throughput Screening Assays for
;    TITLE OF INVENTION:  Modulators of STAT4 and STAT6 Activity
;    NUMBER OF SEQUENCES:  64
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend and Crew LLP
;      STREET:  Two Embarcadero Center, Eighth Floor
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-3834
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/053,003
;      FILING DATE:  31-MAR-1998
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Smith, Timothy L.
;      REGISTRATION NUMBER:  35,367
;      REFERENCE/DOCKET NUMBER:  018781-000800US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 576-0200
;      TELEFAX:  (415) 576-0300
;  INFORMATION FOR SEQ ID NO:  40:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  201 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FEATURE:
;      NAME/KEY:  Modified-site
;      LOCATION:  1..97
;      OTHER INFORMATION:  /product= "OTHER"
;      OTHER INFORMATION:  /note= "Gly at positions 1-97 may be
;      OTHER INFORMATION:  present or absent"
;    FEATURE:
;      NAME/KEY:  Modified-site
;      LOCATION:  105..201
;      OTHER INFORMATION:  /product= "OTHER"
;      OTHER INFORMATION:  /note= "Gly at positions 105-201 may be
;      OTHER INFORMATION:  present or absent"
US-09-053-003-40

  Query Match             9.7%;  Score 209.5;  DB 4;  Length 201;
  Best Local Similarity   40.7%;  Pred. No. 2.9e-09;
  Matches   55;  Conservative    0;  Mismatches   63;  Indels   17;  Gaps    2;

Qy      6 GGGRCCPEQEEGESAAGGSGAGGDSAIEQGGQGSALAPSPVSGVRREGARGGGRGRGRWK 65
          |||         |   || | ||     || |            |    |   ||| |  |
Db     17 GGG---------GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 68

Qy     66 QAGRGGGVCGRGRGRGRGRGRGRGRGRGRPPSGGSGLGGDGGGCGGGGSGGGGAPRRE 125
           | |||  | | | | | | | | | | | |  | ||  | ||  ||| |||| ||||
Db     69 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGPGGGGGGGGGGGGGGGGGGGGGG---- 124

Qy    126 PVPFPSGSAGPGPRG 140
                |  |  |  |
Db    125 -----GGGGGGGGG 134
```

```
RESULT 9
US-09-054-281-22
; Sequence 22, Application US/09054281
; Patent No. 6444421
;  GENERAL INFORMATION:
;    APPLICANT:  Chung, Jay H.
;    TITLE OF INVENTION:  Methods for Detecting Intermolecular
;    TITLE OF INVENTION:  Interactions In Vivo and In Vitro
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend and Crew LLP
;      STREET:  Two Embarcadero Center, Eighth Floor
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-3834
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/054,281
;      FILING DATE:  02-APR-1998
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/080,234
;      FILING DATE:  03-APR-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/826,622
;      FILING DATE:  03-APR-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Smith, Timothy L.
;      REGISTRATION NUMBER:  35,367
;      REFERENCE/DOCKET NUMBER:  015280-295200US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 576-0200
;      TELEFAX:  (415) 576-0300
;  INFORMATION FOR SEQ ID NO:  22:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  201 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FEATURE:
;      NAME/KEY:  Modified-site
;      LOCATION:  1..97
;      OTHER INFORMATION:  /product= "OTHER"
;      OTHER INFORMATION:  /note= "Gly at positions 1-97 may be
;      OTHER INFORMATION:  present or absent"
;    FEATURE:
;      NAME/KEY:  Modified-site
;      LOCATION:  105..201
;      OTHER INFORMATION:  /product= "OTHER"
;      OTHER INFORMATION:  /note= "Gly at positions 105-201 may be
;      OTHER INFORMATION:  present or absent"
US-09-054-281-22

  Query Match             9.7%;  Score 209.5;  DB 4;  Length 201;
  Best Local Similarity   40.7%;  Pred. No. 2.9e-09;
  Matches   55;  Conservative    0;  Mismatches   63;  Indels   17;  Gaps    2;

Qy      6 GGGRCCPEQEEGESAAGGSGAGGDSAIEQGGQGSALAPSPVSGVRREGARGGGRGRGRWK 65
          |||         |   || | ||     || |            |    |   ||| |  |
Db     17 GGG---------GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 68

Qy     66 QAGRGGGVCGRGRGRGRGRGRGRGRGRGRPPSGGSGLGGDGGGCGGGGSGGGGAPRRE 125
           | |||  | | | | | | | | | | | |  | ||  | ||  ||| |||| ||||
Db     69 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGPGGGGGGGGGGGGGGGGGGGGGG---- 124

Qy    126 PVPFPSGSAGPGPRG 140
                |  |  |  |
```

```
Db    125 -----GGGGGGGGGG 134


RESULT 10
US-09-478-948-6
; Sequence 6, Application US/09478948
; Patent No. 6465258
; GENERAL INFORMATION:
;  APPLICANT: Shan, Bei
;  APPLICANT:  Okamoto, Arthur Y.
;  APPLICANT:  Tularik Inc.
;  TITLE OF INVENTION: FXR Receptor-Mediated Modulation of Cholesterol
;  TITLE OF INVENTION:  Metabolism
;  FILE REFERENCE: 018781-001310US
;  CURRENT APPLICATION NUMBER: US/09/478,948
;  CURRENT FILING DATE:  2000-01-06
;  PRIOR APPLICATION NUMBER: US 60/115,249
;  PRIOR FILING DATE: 1999-01-07
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
;   LENGTH: 201
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:flexible linker
;   NAME/KEY: MOD_RES
;   LOCATION: (1)..(97)
;   OTHER INFORMATION: Gly at positions 1-97 may be present or absent
;   NAME/KEY: MOD_RES
;   LOCATION: (105)..(201)
;   OTHER INFORMATION: Gly at positions 105-201 may be present or absent
US-09-478-948-6

  Query Match             9.7%;  Score 209.5;  DB 4;  Length 201;
  Best Local Similarity   40.7%;  Pred. No. 2.9e-09;
  Matches   55;  Conservative    0;  Mismatches   63;  Indels   17;  Gaps    2;

Qy       6 GGGRCCPEQEEGESAAGGSGAGGDSAIEQGGQGSALAPSPVSGVRREGARGGGRGRGRWK 65
           |||         |    || | ||       || |         |    |  ||| | |
Db      17 GGG--------GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 68

Qy      66 QAGRGGGVCGRGRGRGRGRGRGRGRGRGRPPSGGSGLGGDGGGCGGGGSGGGGAPRRE 125
           | |||  | | | | | | | | | | |  | || | || ||| |||| ||||
Db      69 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGPGGGGGGGGGGGGGGGGGGGGGG---- 124

Qy     126 PVPFPSGSAGPGPRG 140
                 |  | |  |
Db     125 -----GGGGGGGGG 134


RESULT 11
US-08-806-029-9
; Sequence 9, Application US/08806029
; Patent No. 6380154
;  GENERAL INFORMATION:
;    APPLICANT:  Cappello, Joseph
;    APPLICANT:  Stedronsky, Erwin R.
;    TITLE OF INVENTION:  Synthetic Proteins for in vivo Drug
;    TITLE OF INVENTION:  Delivery and Tissue Augmentation
;    NUMBER OF SEQUENCES:  36
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert
;      STREET:  Four Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  94111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/806,029
;      FILING DATE:  24-FEB-1997
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/212,237
;      FILING DATE:  11-MAR-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Trecartin, Richard F.
;      REGISTRATION NUMBER:  31,801
;      REFERENCE/DOCKET NUMBER:  A-58847-2/RFT/MTK
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1136 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-806-029-9

  Query Match             9.1%;  Score 196.5;  DB 4;  Length 1136;
  Best Local Similarity   33.0%;  Pred. No. 2.9e-07;
  Matches   59;  Conservative   10;  Mismatches   95;  Indels   15;  Gaps    5;

Qy       3 AHPGGGRCCPEQEEGESAAGGSGAGGDSAIEQG-GQGSALAPSPVSGVRREGARGGGRGR 61
           |  | |          |  |  |||||  |    | | |:             ||       | | |
Db     965 AGAGSGAGAGSGAAGYGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGA 1024

Qy      62 GRWKQAGRGGGVCGRGRGRGRGRGRGRGRGRGRPPSGGSGLGGDGGGCGGGGSGGGGA 121
           |    || | |  | | | | | | | | | | | | |     |:|  | | | | |   | ||
Db    1025 GAGSGAGAGSGAAGYGAGAGSGAGAGSGAGAGSGAGAGSGAG-AGSGAGAGSGAGAGSGA 1083

Qy     122 PRREPVPFPSGSAGPGPRGPRATESGKRMDCPALPPGWKKEEVIRKSGLSAGKSDVYYF 180
                   ||:|| |      : :|      |: ||        :  ||||:   |::
Db    1084 GAGSGAGAGSGAAGYG--AGAGSGAGAGSGAGAMDPG--------RYQLSAGR---YHY 1129


RESULT 12
US-07-609-716-31
; Sequence 31, Application US/07609716
; Patent No. 5514581
;  GENERAL INFORMATION:
;    APPLICANT:  Ferrari, Franco A.
;    APPLICANT:  Cappello, Joseph
;    TITLE OF INVENTION:  Functional Recombinantly Prepared
;    TITLE OF INVENTION:  Synthetic Protein Polymer
;    NUMBER OF SEQUENCES:  118
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert
;      STREET:  Four Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  US
;      ZIP:  94111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/609,716
;      FILING DATE:  06-NOV-1990
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rowland, Bertram I
;      REGISTRATION NUMBER:  20015
;      REFERENCE/DOCKET NUMBER:  A-55186-3/BIR
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-781-1989
;      TELEFAX:  415-398-3249
```

```
;  INFORMATION FOR SEQ ID NO:  31:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1177 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-609-716-31

  Query Match             9.1%;  Score 196.5;  DB 1;  Length 1177;
  Best Local Similarity   33.0%;  Pred. No. 3.1e-07;
  Matches   59;  Conservative   10;  Mismatches   95;  Indels   15;  Gaps    5;

Qy       3 AHPGGGRCCPEQEEGESAAGGSGAGGDSAIEQG-GQGSALAPSPVSGVRREGARGGGRGR 61
           |  | |         |  |  |||||  |    | | |:       ||       | | |
Db    1006 AGAGSGAGAGSGAAGYGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGA 1065

Qy      62 GRWKQAGRGGGVCGRGRGRGRGRGRGRGRGRGRPPSGGSGLGGDGGGCGGGGSGGGGA 121
           |    || | |  | | | | | | | | | | |     |:|  | | | | |   | ||
Db    1066 GAGSGAGAGSGAAGYGAGAGSGAGAGSGAGAGSGAGAGSGAG-AGSGAGAGSGAGAGSGA 1124

Qy     122 PRREPVPFPSGSAGPGPRGPRATESGKRMDCPALPPGWKKEEVIRKSGLSAGKSDVYYF 180
                 ||:|| |      : :|      |: ||        :  ||||:   |::
Db    1125 GAGSGAGAGSGAAGYG--AGAGSGAGAGSGAGAMDPG--------RYQLSAGR---YHY 1170


RESULT 13
US-08-175-155-29
; Sequence 29, Application US/08175155
; Patent No. 5641648
;  GENERAL INFORMATION:
;    APPLICANT:  Ferrari, Franco A.
;    APPLICANT:  Cappello, Joseph
;    APPLICANT:  Crissman, John W.
;    APPLICANT:  Dorman, Mary A.
;    TITLE OF INVENTION:  Methods for Preparing Synthetic
;    TITLE OF INVENTION:  Repetitive DNA
;    NUMBER OF SEQUENCES:  69
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert
;      STREET:  Four Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  US
;      ZIP:  94111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/175,155
;      FILING DATE:  29-DEC-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rowland, Bertram I.
;      REGISTRATION NUMBER:  20015
;      REFERENCE/DOCKET NUMBER:  A-55186-5/BIR
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-781-1989
;      TELEFAX:  415-398-3249
;  INFORMATION FOR SEQ ID NO:  29:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1177 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-175-155-29

  Query Match             9.1%;  Score 196.5;  DB 1;  Length 1177;
  Best Local Similarity   33.0%;  Pred. No. 3.1e-07;
  Matches   59;  Conservative   10;  Mismatches   95;  Indels   15;  Gaps    5;

Qy       3 AHPGGGRCCPEQEEGESAAGGSGAGGDSAIEQG-GQGSALAPSPVSGVRREGARGGGRGR 61
           |  | |         |  |  |||||  |    | | |:       ||       | | |
Db    1006 AGAGSGAGAGSGAAGYGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGA 1065

Qy      62 GRWKQAGRGGGVCGRGRGRGRGRGRGRGRGRGRPPSGGSGLGGDGGGCGGGGSGGGGA 121
           |    || | |  | | | | | | | | | | |     |:|  | | | | |   | ||
Db    1066 GAGSGAGAGSGAAGYGAGAGSGAGAGSGAGAGSGAG-AGSGAGAGSGAGAGSGA 1124

Qy     122 PRREPVPFPSGSAGPGPRGPRATESGKRMDCPALPPGWKKEEVIRKSGLSAGKSDVYYF 180
                 ||:|| |      : :|      |: ||        :  ||||:   |::
Db    1125 GAGSGAGAGSGAAGYG--AGAGSGAGAGSGAGAMDPG--------RYQLSAGR---YHY 1170


RESULT 14
US-08-477-509B-64
; Sequence 64, Application US/08477509B
; Patent No. 5770697
;  GENERAL INFORMATION:
;    APPLICANT:  Ferrari, Franco A
;    APPLICANT:  Cappello, Joseph
;    APPLICANT:  Crissman, John w
;    APPLICANT:  Dorman, Mary A
;    TITLE OF INVENTION:  No. 5770697el Peptides Comprising Repetitive
;    TITLE OF INVENTION:  Units of Amino Acids and DNA Sequences Encoding the Same
;    NUMBER OF SEQUENCES:  112
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert
;      STREET:  Four Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  US
;      ZIP:  94111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/477,509B
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/175,155
;      FILING DATE:  29-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/053,049
;      FILING DATE:  22-APR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/114,618
;      FILING DATE:  29-OCT-1987
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 06/927,258
;      FILING DATE:  04-NOV-1986
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Trecartin, Richard F.
;      REGISTRATION NUMBER:  31,801
;      REFERENCE/DOCKET NUMBER:  A-55186-7/RFT/MTK
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-781-1989
;      TELEFAX:  415-398-3249
;  INFORMATION FOR SEQ ID NO:  64:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1177 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-477-509B-64

  Query Match             9.1%;  Score 196.5;  DB 1;  Length 1177;
  Best Local Similarity   33.0%;  Pred. No. 3.1e-07;
  Matches   59;  Conservative   10;  Mismatches   95;  Indels   15;  Gaps    5;
```

```
Qy       3 AHPGGGRCCPEQEEGESAAGGSGAGGDSAIEQG-GQGSALAPSPVSGVRREGARGGGRGR 61
           |  | |         |  |  |||||  |    | | |:       ||       | | |
Db    1006 AGAGSGAGAGSGAAGYGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGA 1065

Qy      62 GRWKQAGRGGGVCGRGRGRGRGRGRGRGRGRGRPPSGGSGLGGDGGGCGGGGSGGGGA 121
           |    || | |  | | | | | | | | | | | |     |:|  | | | | |   | ||
Db    1066 GAGSGAGAGSGAAGYGAGAGSGAGAGSGAGAGSGAGAGSGAG-AGSGAGAGSGAGAGSGA 1124

Qy     122 PRREPVPFPSGSAGPGPRGPRATESGKRMDCPALPPGWKKEEVIRKSGLSAGKSDVYYF 180
                 ||:|| |      : :|      |: ||        :  ||||:   |::
Db    1125 GAGSGAGAGSGAAGYG--AGAGSGAGAGSGAGAMDPG--------RYQLSAGR---YHY 1170


RESULT 15
US-08-707-237A-35
; Sequence 35, Application US/08707237A
; Patent No. 5830713
;  GENERAL INFORMATION:
;    APPLICANT:  Ferrari, Franco A.
;    APPLICANT:  Capello, Joseph
;    APPLICANT:  Crissman, John W.
;    APPLICANT:  Dorman, Mary A.
;    TITLE OF INVENTION:  METHODS FOR PREPARING SYNTHETIC
;    TITLE OF INVENTION:  REPETITIVE DNA
;    NUMBER OF SEQUENCES:  108
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert
;      STREET:  Four Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/707,237A
;      FILING DATE:  03-SEP-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/175,155
;      FILING DATE:  29-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/053,049
;      FILING DATE:  22-APR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/609,716
;      FILING DATE:  06-NOV-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/269,429
;      FILING DATE:  09-NOV-1988
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/114,618
;      FILING DATE:  29-OCT-1987
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 06/927,258
;      FILING DATE:  04-NOV-1986
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Trecartin, Richard F.
;      REGISTRATION NUMBER:  31,801
;      REFERENCE/DOCKET NUMBER:  A-55186-10/WHD
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  35:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1177 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-707-237A-35

  Query Match             9.1%;  Score 196.5;  DB 2;  Length 1177;
  Best Local Similarity   33.0%;  Pred. No. 3.1e-07;
  Matches   59;  Conservative   10;  Mismatches   95;  Indels   15;  Gaps    5;

Qy       3 AHPGGGRCCPEQEEGESAAGGSGAGGDSAIEQG-GQGSALAPSPVSGVRREGARGGGRGR 61
           |  | |         |  |  |||||  |    | | |:       ||       | | |
Db    1006 AGAGSGAGAGSGAAGYGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGA 1065

Qy      62 GRWKQAGRGGGVCGRGRGRGRGRGRGRGRGRGRPPSGGSGLGGDGGGCGGGGSGGGGA 121
           |    || | |  | | | | | | | | | | | |     |:|  | | | | |   | ||
Db    1066 GAGSGAGAGSGAAGYGAGAGSGAGAGSGAGAGSGAGAGSGAG-AGSGAGAGSGAGAGSGA 1124

Qy     122 PRREPVPFPSGSAGPGPRGPRATESGKRMDCPALPPGWKKEEVIRKSGLSAGKSDVYYF 180
                 ||:|| |      : :|      |: ||        :  ||||:   |::
Db    1125 GAGSGAGAGSGAAGYG--AGAGSGAGAGSGAGAMDPG--------RYQLSAGR---YHY 1170


Search completed: March 12, 2003, 09:15:32
Job time : 29.5423 secs
```

GenCore version 5.1.4_p5_4578

OM protein - protein search, using sw model

Run on: March 12, 2003, 00:59:01 ; Search time 59.8484 Seconds
(without alignments)
915.078 Million cell updates/sec

Title: US-09-554-414B-2
Perfect score: 2167
Sequence: 1 MRAHPGGGRCCPEQEEGESA.....LSRAADTEEMDIEMDSGDEA 411

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2167	100.0	411	20	AAY14197	Human DNA demethyl
2	2167	100.0	411	22	AAB99915	Human protein sequ
3	2167	100.0	411	22	AAG64314	Human protein #1.
4	2167	100.0	411	22	AAG64844	Heart muscle cell
5	2084.5	96.2	414	20	AAY14199	Mouse DNA demethyl
6	1445	66.7	281	20	AAY73829	Human prostate tum
7	1445	66.7	281	20	AAY48439	Human prostate can
8	1344	62.0	263	19	AAW74980	Human secreted pro
9	1059	48.9	282	23	AAE22578	Xenopus laevis MBD
10	1056.5	48.8	303	23	AAE22577	Xenopus laevis MBD
11	1030	47.5	291	20	AAY14198	Human DNA demethyl
12	1021.5	47.1	285	20	AAY14200	Mouse DNA demethyl
13	1013	46.7	200	20	AAY07107	Colon cancer assoc
14	753	34.7	223	22	AAB92997	Human protein sequ
15	739	34.1	218	22	AAB93733	Human protein sequ
16	452.5	20.9	314	22	ABB66035	Drosophila melanog
17	444	20.5	339	23	AAE22575	Drosophila dMBD-li
18	400.5	18.5	226	22	ABB63747	Drosophila melanog
19	400.5	18.5	226	23	AAE22576	Drosophila dMBD-li
20	341	15.7	64	23	AAE22569	Human MBD2 methyl
21	280.5	12.9	63	23	AAE22570	Xenopus laevis MBD
22	218.5	10.1	610	22	ABB67896	Drosophila melanog
23	214.5	9.9	574	21	AAY28900	Regulator of fibro
24	214.5	9.9	590	21	AAB21008	Human nucleic acid
25	214.5	9.9	629	21	AAY28899	Regulator of fibro
26	214.5	9.9	630	21	AAY28898	Regulator of fibro
27	214	9.9	641	20	AAY28843	Epstein Barr Virus
28	214	9.9	641	21	AAY95856	Epstein Barr virus
29	214	9.9	641	22	AAB62332	EBV tethering prot
30	210	9.7	235	17	AAW05704	Glycine-rich repea
31	210	9.7	235	19	AAW79126	Epstein Barr Virus
32	210	9.7	261	19	AAW79137	FLGA Gly-ala inser
33	209	9.6	64	23	AAE22567	Human MBD1 methyl
34	207.5	9.6	174	21	AAG39747	Arabidopsis thalia
35	203.5	9.4	357	22	ABB29912	Peptide #2563 enco
36	203.5	9.4	357	22	ABB35090	Peptide #2596 enco
37	203.5	9.4	357	22	ABB20509	Protein #2508 enco
38	203.5	9.4	357	22	AAM55912	Human brain expres
39	203.5	9.4	357	22	AAM68282	Human bone marrow
40	203.5	9.4	357	22	AAM16105	Peptide #2539 enco
41	203.5	9.4	357	22	AAM28596	Peptide #2633 enco
42	203.5	9.4	357	22	AAM03832	Peptide #2514 enco
43	203.5	9.4	357	23	ABG37823	Human peptide enco
44	199	9.2	403	17	AAR88587	EBV nuclear antige
45	197	9.1	344	22	ABB64729	Drosophila melanog

ALIGNMENTS

RESULT 1
AAY14197
ID AAY14197 standard; Protein; 411 AA.
XX
AC AAY14197;
XX
DT 28-JUL-1999 (first entry)
XX
DE Human DNA demethylase, dMTase1, protien sequence.
XX
KW DNA demethylase; dMTase1; dMTase2; human; cancer cell; beta-thalassemia;
KW DNA methylation pattern; methylation inhibitor; gene therapy; diagnosis;
KW genetic defect correction; sickle cell anaemia.
XX
OS Homo sapiens.
XX
PN WO9924583-A1.
XX
PD 20-MAY-1999.
XX
PF 12-NOV-1998; 98WO-CA01059.
XX
PR 11-MAY-1998; 98CA-2230991.
PR 12-NOV-1997; 97CA-2220805.
XX
PA (UYMC-) UNIV MCGILL.
XX
PI Bhattacharya S, Ramchandani S, Szyf M;
XX
DR WPI; 1999-347283/29.
DR N-PSDB; AAX61218.
XX
PT Human and murine DNA demethylases useful in the diagnosis of cancer

```
XX
PS   Disclosure; Fig 9c; 114pp; English.
XX
CC   This sequence is the human DNA demethylase, designated dMTase1, of
CC   the invention. The DNA demethylase is overexpressed in cancer cells.
CC   Expression of the demethylase cDNA is useful to alter DNA methylation
CC   patterns of DNA in vitro in cells or in vivo in humans, animals and
CC   plants. The cDNA is in antisense orientation to inhibit demethylase in
CC   cancer cells for therapeutic purposes. The demethylase is used to alter
CC   the differentiation state and to generate stem cells for therapeutics,
CC   cells for animal cloning and to improve expression of foreign genes. The
CC   cDNA can also be used for recombinant production of large amounts of the
CC   demethylase. The protein can be used to raise antibodies against
CC   demethylase. It can also be used for high throughput screening of
CC   demethylase inhibitors, and for obtaining the x-ray crystal structure.
CC   The demethylase cDNA and protein are also useful for changing the state
CC   of differentiation of a cell to allow gene therapy, stem cell selection
CC   or cell cloning; or for inhibiting methylation in cancer cells using
CC   vector mediated gene therapy. Antagonists or inhibitors of the
CC   demethylase can be used to manufacture medicaments for cancer treatment,
CC   for restoring an aberrant methylation patterns or changing methylation
CC   patterns in patient DNA. Change of the methylation pattern activates a
CC   silent gene permitting the correction of a genetic defect, such as
CC   beta-thalassemia or sickle cell anaemia. The cDNA can be used as a
CC   template to design antisense oligonucleotides and ribozymes. The cDNA can
CC   also be used in two-hybrid systems in yeast to identify proteins
CC   interacting with demethylase. Determining the level of DNA methylase
CC   expression in a cell can be used as a method for diagnosis of cancer,
CC   where overexpression is indicative of cancer cells.
XX
SQ   Sequence   411 AA;

  Query Match             100.0%;  Score 2167;  DB 20;  Length 411;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-165;
  Matches  411;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MRAHPGGGRCCPEQEEGESAAGGSGAGGDSAIEQGGQGSALAPSPVSGVRREGARGGGRG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MRAHPGGGRCCPEQEEGESAAGGSGAGGDSAIEQGGQGSALAPSPVSGVRREGARGGGRG 60

Qy      61 RGRWKQAGRGGGVCGRGRGRGRGRGRGRGRGRGRGRPPSGGSGLGGDGGGCGGGGSGGGG 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 RGRWKQAGRGGGVCGRGRGRGRGRGRGRGRGRGRGRPPSGGSGLGGDGGGCGGGGSGGGG 120

Qy     121 APRREPVPFPSGSAGPGPRGPRATESGKRMDCPALPPGWKKEEVIRKSGLSAGKSDVYYF 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 APRREPVPFPSGSAGPGPRGPRATESGKRMDCPALPPGWKKEEVIRKSGLSAGKSDVYYF 180

Qy     181 SPSGKKFRSKPQLARYLGNTVDLSSFDFRTGKMMPSKLQKNKQRLRNDPLNQNKGKPDLN 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 SPSGKKFRSKPQLARYLGNTVDLSSFDFRTGKMMPSKLQKNKQRLRNDPLNQNKGKPDLN 240

Qy     241 TTLPIRQTASIFKQPVTKVTNHPSNKVKSDPQRMNEQPRQLFWEKRLQGLSASDVTEQII 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 TTLPIRQTASIFKQPVTKVTNHPSNKVKSDPQRMNEQPRQLFWEKRLQGLSASDVTEQII 300

Qy     301 KTMELPKGLQGVGPGSNDETLLSAVASALHTSSAPITGQVSAAVEKNPAVWLNTSQPLCK 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 KTMELPKGLQGVGPGSNDETLLSAVASALHTSSAPITGQVSAAVEKNPAVWLNTSQPLCK 360

Qy     361 AFIVTDEDIRKQEERVQQVRKKLEEALMADILSRAADTEEMDIEMDSGDEA 411
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 AFIVTDEDIRKQEERVQQVRKKLEEALMADILSRAADTEEMDIEMDSGDEA 411


RESULT 2
AAB99915
ID   AAB99915 standard; Protein; 411 AA.
XX
AC   AAB99915;
XX
DT   26-SEP-2001  (first entry)
XX
DE   Human protein sequence SEQ ID NO:1.
XX
KW   Differentiation; heart muscle cell; cytokine; transcription factor;
KW   proliferation; surface antigen; heart disease; cardiomyocyte;
KW   bone marrow; umbilical blood cell; heart muscle degeneration;
KW   myocardial infarction.
XX
OS   Homo sapiens.
XX
PN   WO200148150-A1.
XX
PD   05-JUL-2001.
XX
PF   02-NOV-2000; 2000WO-JP07741.
XX
PR   28-DEC-1999;   99JP-0372826.
PR   28-FEB-2000; 2000WO-JP01148.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI   Umezawa A,  Hata J,  Fukuda K,  Ogawa S,  Sakurada K,  Gojo S;
PI   Yamada Y;
XX
DR   WPI; 2001-425655/45.
DR   N-PSDB; AAH44351.
XX
PT   Cells capable of differentiating into cardiomyocytes and originating in
PT   bone marrow or umbilical blood cells for study of cardiomyocyte
PT   differentiation and treatment of heart disease -
XX
PS   Claim 22; Page 84-86; 187pp; Japanese.
XX
CC   The present invention describes cells originating in bone marrow or
CC   umbilical blood cells which are capable of differentiating into
CC   cardiomyocytes. Also described are: (1) cardiomyocytes produced by the
CC   differentiation of the cells; (2) a method for carrying out the
CC   differentiation into cardiomyocytes, regulated by a promotional and/or
CC   inhibitory factor; (3) a method for the differentiation of the cells
CC   into cell types other than cardiomyocytes; (4) drug compositions
CC   promoting the formation of heart muscle and regeneration of heart tissue
CC   which contain the cells; (5) a method for the production of antibodies
CC   which recognise the cells, especially antibodies which recognise a
CC   surface antigen on the cells; (6) a method for screening factors which
CC   promote the proliferation of the cells; (7) a method for immortalising
CC   the cells by expressing telomerase in them; (8) drug compositions for
CC   the treatment of heart disease which contain the immortalised cells; and
CC   (9) cell-free supernatant from the culture of the cells and its use in
CC   promoting their differentiation into cardiomyocytes. The cells are used
CC   in the treatment of diseases involving heart muscle degeneration, such
CC   as myocardial infarction and in the study of cardiomyocyte
CC   differentiation. AAH44351 to AAH44409 and AAB99915 to AAB99935 represent
CC   sequences used in the exemplification of the present invention.
XX
SQ   Sequence   411 AA;

  Query Match             100.0%;  Score 2167;  DB 22;  Length 411;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-165;
  Matches  411;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MRAHPGGGRCCPEQEEGESAAGGSGAGGDSAIEQGGQGSALAPSPVSGVRREGARGGGRG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MRAHPGGGRCCPEQEEGESAAGGSGAGGDSAIEQGGQGSALAPSPVSGVRREGARGGGRG 60

Qy      61 RGRWKQAGRGGGVCGRGRGRGRGRGRGRGRGRGRPPSGGSGLGGDGGGCGGGGSGGGG 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 RGRWKQAGRGGGVCGRGRGRGRGRGRGRGRGRGRPPSGGSGLGGDGGGCGGGGSGGGG 120

Qy     121 APRREPVPFPSGSAGPGPRGPRATESGKRMDCPALPPGWKKEEVIRKSGLSAGKSDVYYF 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 APRREPVPFPSGSAGPGPRGPRATESGKRMDCPALPPGWKKEEVIRKSGLSAGKSDVYYF 180

Qy     181 SPSGKKFRSKPQLARYLGNTVDLSSFDFRTGKMMPSKLQKNKQRLRNDPLNQNKGKPDLN 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    181 SPSGKKFRSKPQLARYLGNTVDLSSFDFRTGKMMPSKLQKNKQRLRNDPLNQNKGKPDLN 240

Qy    241 TTLPIRQTASIFKQPVTKVTNHPSNKVKSDPQRMNEQPRQLFWEKRLQGLSASDVTEQII 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TTLPIRQTASIFKQPVTKVTNHPSNKVKSDPQRMNEQPRQLFWEKRLQGLSASDVTEQII 300

Qy    301 KTMELPKGLQGVGPGSNDETLLSAVASALHTSSAPITGQVSAAVEKNPAVWLNTSQPLCK 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 KTMELPKGLQGVGPGSNDETLLSAVASALHTSSAPITGQVSAAVEKNPAVWLNTSQPLCK 360

Qy    361 AFIVTDEDIRKQEERVQQVRKKLEEALMADILSRAADTEEMDIEMDSGDEA 411
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 AFIVTDEDIRKQEERVQQVRKKLEEALMADILSRAADTEEMDIEMDSGDEA 411


RESULT 3
AAG64314
ID   AAG64314 standard; Protein; 411 AA.
XX
AC   AAG64314;
XX
DT   24-SEP-2001  (first entry)
XX
DE   Human protein #1.
XX
KW   Angiogenesis; cardiant; cell differentiating agent; bone marrow;
KW   heart muscle cell; heart disease; human.
XX
OS   Homo sapiens.
XX
PN   WO200148149-A1.
XX
PD   05-JUL-2001.
XX
PF   28-FEB-2000; 2000WO-JP01148.
XX
PR   28-DEC-1999;   99JP-0372826.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI   Umezawa A,  Hata J,  Fukuda K,  Ogawa S,  Sakurada K;
XX
DR   WPI; 2001-418252/44.
DR   N-PSDB; AAH49586.
XX
PT   New adult bone marrow-originated cells capable of differentiating into
PT   heart muscle cells, applicable as remedies for various heart diseases
PT   particularly with damaged heart muscle accompanying degeneration  -
XX
PS   Claim 12; Pages 55-57; 158pp; Japanese.
XX
CC   The present invention relates to cells isolated from bone marrow, which
CC   are capable of at least differentiating into heart muscle cells. The
CC   cells are applicable as remedies for various heart diseases particularly
CC   with damaged heart muscle accompanying degeneration. The present sequence
CC   was used to illustrate the present invention.
XX
SQ   Sequence   411 AA;

  Query Match           100.0%;  Score 2167;  DB 22;  Length 411;
  Best Local Similarity  100.0%;  Pred. No. 1.1e-165;
  Matches  411;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 MRAHPGGGRCCPEQEEGESAAGGSGAGGDSAIEQGGQGSALAPSPVSGVRREGARGGGRG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MRAHPGGGRCCPEQEEGESAAGGSGAGGDSAIEQGGQGSALAPSPVSGVRREGARGGGRG 60

Qy     61 RGRWKQAGRGGGVCGRGRGRGRGRGRGRGRGRGRPPSGGSGLGGDGGGCGGGGSGGGG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 RGRWKQAGRGGGVCGRGRGRGRGRGRGRGRGRGRPPSGGSGLGGDGGGCGGGGSGGGG 120

Qy    121 APRREPVPFPSGSAGPGPRGPRATESGKRMDCPALPPGWKKEEVIRKSGLSAGKSDVYYF 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 APRREPVPFPSGSAGPGPRGPRATESGKRMDCPALPPGWKKEEVIRKSGLSAGKSDVYYF 180

Qy    181 SPSGKKFRSKPQLARYLGNTVDLSSFDFRTGKMMPSKLQKNKQRLRNDPLNQNKGKPDLN 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 SPSGKKFRSKPQLARYLGNTVDLSSFDFRTGKMMPSKLQKNKQRLRNDPLNQNKGKPDLN 240

Qy    241 TTLPIRQTASIFKQPVTKVTNHPSNKVKSDPQRMNEQPRQLFWEKRLQGLSASDVTEQII 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TTLPIRQTASIFKQPVTKVTNHPSNKVKSDPQRMNEQPRQLFWEKRLQGLSASDVTEQII 300

Qy    301 KTMELPKGLQGVGPGSNDETLLSAVASALHTSSAPITGQVSAAVEKNPAVWLNTSQPLCK 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 KTMELPKGLQGVGPGSNDETLLSAVASALHTSSAPITGQVSAAVEKNPAVWLNTSQPLCK 360

Qy    361 AFIVTDEDIRKQEERVQQVRKKLEEALMADILSRAADTEEMDIEMDSGDEA 411
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 AFIVTDEDIRKQEERVQQVRKKLEEALMADILSRAADTEEMDIEMDSGDEA 411


RESULT 4
AAG64844
ID   AAG64844 standard; Protein; 411 AA.
XX
AC   AAG64844;
XX
DT   21-SEP-2001  (first entry)
XX
DE   Heart muscle cell differentiation related protein SEQ ID NO: 1.
XX
KW   Heart muscle cell; human; cell differentiation; heart disease.
XX
OS   Homo sapiens.
XX
PN   WO200148151-A1.
XX
PD   05-JUL-2001.
XX
PF   27-DEC-2000; 2000WO-JP09323.
XX
PR   28-DEC-1999;   99JP-0372826.
PR   28-FEB-2000; 2000WO-JP01148.
PR   02-NOV-2000; 2000WO-JP07741.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI   Umezawa A,  Hata J,  Fukuda K,  Ogawa S,  Sakurada K,  Gojo S;
PI   Yamada Y;
XX
DR   WPI; 2001-425656/45.
DR   N-PSDB; AAH48220.
XX
PT   Cells capable of differentiating into cardiomyocytes and originating in
PT   bone marrow or umbilical blood cells for study of cardiomyocyte
PT   differentiation and treatment of heart disease  -
XX
PS   Claim 28; Page 90-92; 183pp; Japanese.
XX
CC   The present invention provides cells originating in the human bone marrow
CC   or umbilical blood cells which are capable of differentiating into
CC   cardiomyocytes. These cells are useful in the treatment of diseases
CC   involving heart muscle degeneration, such as myocardial infarction, and
CC   the study of cardiomyocyte differentiation. The present sequence is
CC   a protein described in the exemplification of the invention.
XX
SQ   Sequence   411 AA;

  Query Match           100.0%;  Score 2167;  DB 22;  Length 411;
  Best Local Similarity  100.0%;  Pred. No. 1.1e-165;
  Matches  411;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 MRAHPGGGRCCPEQEEGESAAGGSGAGGDSAIEQGGQGSALAPSPVSGVRREGARGGGRG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MRAHPGGGRCCPEQEEGESAAGGSGAGGDSAIEQGGQGSALAPSPVSGVRREGARGGGRG 60
```

```
Qy      61 RGRWKQAGRGGGVCGRGRGRGRGRGRGRGRGRGRPPSGGSGLGGDGGGCGGGGSGGGG 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 RGRWKQAGRGGGVCGRGRGRGRGRGRGRGRGRGRPPSGGSGLGGDGGGCGGGGSGGGG 120

Qy     121 APRREPVPFPSGSAGPGPRGPRATESGKRMDCPALPPGWKKEEVIRKSGLSAGKSDVYYF 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 APRREPVPFPSGSAGPGPRGPRATESGKRMDCPALPPGWKKEEVIRKSGLSAGKSDVYYF 180

Qy     181 SPSGKKFRSKPQLARYLGNTVDLSSFDFRTGKMMPSKLQKNKQRLRNDPLNQNKGKPDLN 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 SPSGKKFRSKPQLARYLGNTVDLSSFDFRTGKMMPSKLQKNKQRLRNDPLNQNKGKPDLN 240

Qy     241 TTLPIRQTASIFKQPVTKVTNHPSNKVKSDPQRMNEQPRQLFWEKRLQGLSASDVTEQII 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 TTLPIRQTASIFKQPVTKVTNHPSNKVKSDPQRMNEQPRQLFWEKRLQGLSASDVTEQII 300

Qy     301 KTMELPKGLQGVGPGSNDETLLSAVASALHTSSAPITGQVSAAVEKNPAVWLNTSQPLCK 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 KTMELPKGLQGVGPGSNDETLLSAVASALHTSSAPITGQVSAAVEKNPAVWLNTSQPLCK 360

Qy     361 AFIVTDEDIRKQEERVQQVRKKLEEALMADILSRAADTEEMDIEMDSGDEA 411
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 AFIVTDEDIRKQEERVQQVRKKLEEALMADILSRAADTEEMDIEMDSGDEA 411
```

RESULT 5
AAY14199
ID AAY14199 standard; Protein; 414 AA.
XX
AC AAY14199;
XX
DT 28-JUL-1999 (first entry)
XX
DE Mouse DNA demethylase, dMTase1, protien sequence.
XX
KW DNA demethylase; dMTase1; dMTase2; mouse; cancer cell; beta-thalassemia;
KW DNA methylation pattern; methylation inhibitor; gene therapy; diagnosis;
KW genetic defect correction; sickle cell anaemia.
XX
OS Mus sp.
XX
PN WO9924583-A1.
XX
PD 20-MAY-1999.
XX
PF 12-NOV-1998; 98WO-CA01059.
XX
PR 11-MAY-1998; 98CA-2230991.
PR 12-NOV-1997; 97CA-2220805.
XX
PA (UYMC-) UNIV MCGILL.
XX
PI Bhattacharya S, Ramchandani S, Szyf M;
XX
DR WPI; 1999-347283/29.
DR N-PSDB; AAX61220.
XX
PT Human and murine DNA demethylases useful in the diagnosis of cancer
XX
PS Disclosure; Fig 9k; 114pp; English.
XX
CC This sequence is the mouse DNA demethylase, designated dMTase1, of
CC the invention. The DNA demethylase is overexpressed in cancer cells.
CC Expression of the demethylase cDNA is useful to alter DNA methylation
CC patterns of DNA in vitro in cells or in vivo in humans, animals and
CC plants. The cDNA is in antisense orientation to inhibit demethylase in
CC cancer cells for therapeutic purposes. The demethylase is used to alter
CC the differentiation state and to generate stem cells for therapeutics,
CC cells for animal cloning and to improve expression of foreign genes. The
CC cDNA can also be used for recombinant production of large amounts of the
CC demethylase. The protein can be used to raise antibodies against
CC demethylase. It can also be used for high throughput screening of
CC demethylase inhibitors, and for obtaining the x-ray crystal structure.
CC The demethylase cDNA and protein are also useful for changing the state
CC of differentiation of a cell to allow gene therapy, stem cell selection
CC or cell cloning; or for inhibiting methylation in cancer cells using
CC vector mediated gene therapy. Antagonists or inhibitors of the
CC demethylase can be used to manufacture medicaments for cancer treatment,
CC for restoring an aberrant methylation patterns or changing methylation
CC patterns in patient DNA. Change of the methylation pattern activates a
CC silent gene permitting the correction of a genetic defect, such as
CC beta-thalassemia or sickle cell anaemia. The cDNA can be used as a
CC template to design antisense oligonucleotides and ribozymes. The cDNA can
CC also be used in two-hybrid systems in yeast to identify proteins
CC interacting with demethylase. Determining the level of DNA methylase
CC expression in a cell can be used as a method for diagnosis of cancer,
CC where overexpression is indicative of cancer cells.
XX
SQ Sequence 414 AA;

Query Match 96.2%; Score 2084.5; DB 20; Length 414;
Best Local Similarity 96.4%; Pred. No. 4.7e-159;
Matches 399; Conservative 4; Mismatches 8; Indels 3; Gaps 1;

```
Qy       1 MRAHPGGGRCCPEQEEGESAAGGSGAGGDSAIEQGGQGSALAPSPVSGVRREGARGGGRG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db       1 MRAHPGGGRCCPEQEEGESAAGGSGAGGDSAIEQGGQGSALAPSPVSGVRRGGARGGGRG 60

Qy      61 RGRWKQAGRGGGVCGRGRGRGRGRGRGRGRGRGRPPSGGSGLGGDGGGCGGG---GSG 117
           ||||||| |||||||||||||||||||||||||| ||||||||||||  ||   |||
Db      61 RGRWKQAARGGGVCGRGRGRGRGRGRGRGRGRGRPQSGGSGLGGDGGGGAGGCGVGSG 120

Qy     118 GGGAPRREPVPFPSGSAGPGPRGPRATESGKRMDCPALPPGWKKEEVIRKSGLSAGKSDV 177
           || ||||:||||||||:||||||||||||||||||||||||||||||||||||||||||||
Db     121 GGVAPRRDPVPFPSGSSGPGPRGPRATESGKRMDCPALPPGWKKEEVIRKSGLSAGKSDV 180

Qy     178 YYFSPSGKKFRSKPQLARYLGNTVDLSSFDFRTGKMMPSKLQKNKQRLRNDPLNQNKGKP 237
           |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db     181 YYFSPSGKKFRSKPQLARYLGNAVDLSSFDFRTGKMMPSKLQKNKQRLRNDPLNQNKGKP 240

Qy     238 DLNTTLPIRQTASIFKQPVTKVTNHPSNKVKSDPQRMNEQPRQLFWEKRLQGLSASDVTE 297
           |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db     241 DLNTTLPIRQTASIFKQPVTKFTNHPSNKVKSDPQRMNEQPRQLFWEKRLQGLSASDVTE 300

Qy     298 QIIKTMELPKGLQGVGPGSNDETLLSAVASALHTSSAPITGQVSAAVEKNPAVWLNTSQP 357
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 QIIKTMELPKGLQGVGPGSNDETLLSAVASALHTSSAPITGQVSAAVEKNPAVWLNTSQP 360

Qy     358 LCKAFIVTDEDIRKQEERVQQVRKKLEEALMADILSRAADTEEMDIEMDSGDEA 411
           |||||||||||||||||||||||||||||||||||||||:||:|||||||
Db     361 LCKAFIVTDEDIRKQEERVQQVRKKLEEALMADILSRAADTEEVDIDMDSGDEA 414
```

RESULT 6
AAY73829
ID AAY73829 standard; Protein; 281 AA.
XX
AC AAY73829;
XX
DT 14-MAR-2000 (first entry)
XX
DE Human prostate tumor EST fragment derived protein #16.
XX
KW Pancreas; tumor; EST; expressed sequence tag; human; cytostatic;
KW treatment.
XX
OS Homo sapiens.
XX
PN DE19820190-A1.
XX
PD 04-NOV-1999.
XX
PF 28-APR-1998; 98DE-1020190.
XX
PR 28-APR-1998; 98DE-1020190.

```
XX
PA   (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
PI   Rosenthal A,  Specht T,  Hinzmann B,  Schmitt A,  Pilarsky C,  Dahl E;
XX
DR   WPI; 1999-621386/54.
DR   N-PSDB; AAZ52863.
XX
PT   New human nucleic acid sequences from pancreatic tumors, and related
PT   proteins  -
XX
PS   Claim 23; Page 315; 502pp; German.
XX
CC   This invention describes novel polypeptides and their encoding nucleic
CC   acids derived from human pancreatic tumor tissue which have cytostatic
CC   activity. The sequences are also useful in producing pharmaceutical
CC   compositions for treatment of pancreatic tumors. AAY73814-Y74252
CC   represent protein fragments encoded by the human pancreatic tumor cDNA
CC   library derived expressed sequence tag (EST) sequences represented in
CC   AAZ52858-Z53014.
XX
SQ   Sequence   281 AA;

  Query Match             66.7%;  Score 1445;  DB 20;  Length 281;
  Best Local Similarity   100.0%;  Pred. No. 6.5e-108;
  Matches  281;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     131 SGSAGPGPRGPRATESGKRMDCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSK 190
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 SGSAGPGPRGPRATESGKRMDCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSK 60

Qy     191 PQLARYLGNTVDLSSFDFRTGKMMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTAS 250
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 PQLARYLGNTVDLSSFDFRTGKMMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTAS 120

Qy     251 IFKQPVTKVTNHPSNKVKSDPQRMNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQ 310
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 IFKQPVTKVTNHPSNKVKSDPQRMNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQ 180

Qy     311 GVGPGSNDETLLSAVASALHTSSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIR 370
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GVGPGSNDETLLSAVASALHTSSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIR 240

Qy     371 KQEERVQQVRKKLEEALMADILSRAADTEEMDIEMDSGDEA 411
           |||||||||||||||||||||||||||||||||||||||||
Db     241 KQEERVQQVRKKLEEALMADILSRAADTEEMDIEMDSGDEA 281


RESULT 7
AAY48439
ID   AAY48439 standard; Protein; 281 AA.
XX
AC   AAY48439;
XX
DT   08-DEC-1999  (first entry)
XX
DE   Human prostate cancer-associated protein 136.
XX
KW   Expressed sequence tag; EST; prostate; tumor; treatment; gene therapy;
KW   cancer; tissue specificity; human.
XX
OS   Homo sapiens.
XX
PN   DE19811194-A1.
XX
PD   16-SEP-1999.
XX
PF   10-MAR-1998;   98DE-1011194.
XX
PR   10-MAR-1998;   98DE-1011194.
XX
PA   (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
PI   Specht T,  Hinzmann B,  Schmitt A,  Pilarsky C,  Dahl E,  Rosenthal A;
XX
DR   WPI; 1999-519629/44.
DR   N-PSDB; AAZ33533.
XX
PT   New nucleic acid expressed at high level in normal prostatic tissue and
PT   encoded polypeptides, used to treat cancer and screen for therapeutic
PT   agents  -
XX
PS   Claim 22; 176; 194pp; German.
XX
CC   This invention describes novel nucleic acid sequences (A) that are
CC   expressed at high level in normal prostatic tissue. Polypeptides (I)
CC   encoded by (A) are used: (a) for identifying agents for treatment of
CC   prostatic cancer and (b) for therapy of prostate cancer, optionally
CC   where expressed by gene therapy methods. (A) is also used to isolate
CC   full-length genes (for gene therapy) and for recombinant production of
CC   (I), which can be used to raise specific antibodies. (A) are identified
CC   by assembly of ESTs (expressed sequence tags) before these are analyzed
CC   for expression pattern (tissue specificity). This approach eliminates
CC   many of the false results, as regards tissue specificity, associated
CC   with known methods that use single (usually short) ESTs. AAY48304-Y48456
CC   represent peptides encoded by the expressed sequence tags described in
CC   the method of the invention.
XX
SQ   Sequence   281 AA;

  Query Match             66.7%;  Score 1445;  DB 20;  Length 281;
  Best Local Similarity   100.0%;  Pred. No. 6.5e-108;
  Matches  281;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     131 SGSAGPGPRGPRATESGKRMDCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSK 190
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 SGSAGPGPRGPRATESGKRMDCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSK 60

Qy     191 PQLARYLGNTVDLSSFDFRTGKMMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTAS 250
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 PQLARYLGNTVDLSSFDFRTGKMMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTAS 120

Qy     251 IFKQPVTKVTNHPSNKVKSDPQRMNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQ 310
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 IFKQPVTKVTNHPSNKVKSDPQRMNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQ 180

Qy     311 GVGPGSNDETLLSAVASALHTSSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIR 370
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GVGPGSNDETLLSAVASALHTSSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIR 240

Qy     371 KQEERVQQVRKKLEEALMADILSRAADTEEMDIEMDSGDEA 411
           |||||||||||||||||||||||||||||||||||||||||
Db     241 KQEERVQQVRKKLEEALMADILSRAADTEEMDIEMDSGDEA 281


RESULT 8
AAW74980
ID   AAW74980 standard; Protein; 263 AA.
XX
AC   AAW74980;
XX
DT   25-JAN-1999  (first entry)
XX
DE   Human secreted protein encoded by gene 106 clone HT3AM65.
XX
KW   Human; secreted protein; testis; tumour; foetal brain tissue;
KW   fusion protein; cancer; central nervous system; seizure;
KW   diagnosis; neurodegenerative disease.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 263
FT                   /label= unknown
XX
PN   WO9839448-A2.
```

XX
PD 11-SEP-1998.
XX
PF 06-MAR-1998; 98WO-US04493.
XX
PR 02-OCT-1997; 97US-0061060.
PR 07-MAR-1997; 97US-0038621.
PR 07-MAR-1997; 97US-0040161.
PR 07-MAR-1997; 97US-0040162.
PR 07-MAR-1997; 97US-0040163.
PR 07-MAR-1997; 97US-0040333.
PR 07-MAR-1997; 97US-0040334.
PR 07-MAR-1997; 97US-0040336.
PR 07-MAR-1997; 97US-0040626.
PR 11-APR-1997; 97US-0043311.
PR 11-APR-1997; 97US-0043312.
PR 11-APR-1997; 97US-0043313.
PR 11-APR-1997; 97US-0043314.
PR 11-APR-1997; 97US-0043568.
PR 11-APR-1997; 97US-0043569.
PR 11-APR-1997; 97US-0043576.
PR 11-APR-1997; 97US-0043578.
PR 11-APR-1997; 97US-0043580.
PR 11-APR-1997; 97US-0043669.
PR 11-APR-1997; 97US-0043670.
PR 11-APR-1997; 97US-0043671.
PR 11-APR-1997; 97US-0043672.
PR 11-APR-1997; 97US-0043674.
PR 23-MAY-1997; 97US-0047492.
PR 23-MAY-1997; 97US-0047500.
PR 23-MAY-1997; 97US-0047501.
PR 23-MAY-1997; 97US-0047502.
PR 23-MAY-1997; 97US-0047503.
PR 23-MAY-1997; 97US-0047581.
PR 23-MAY-1997; 97US-0047582.
PR 23-MAY-1997; 97US-0047583.
PR 23-MAY-1997; 97US-0047584.
PR 23-MAY-1997; 97US-0047585.
PR 23-MAY-1997; 97US-0047586.
PR 23-MAY-1997; 97US-0047587.
PR 23-MAY-1997; 97US-0047588.
PR 23-MAY-1997; 97US-0047589.
PR 23-MAY-1997; 97US-0047590.
PR 23-MAY-1997; 97US-0047592.
PR 23-MAY-1997; 97US-0047593.
PR 23-MAY-1997; 97US-0047594.
PR 23-MAY-1997; 97US-0047595.
PR 23-MAY-1997; 97US-0047596.
PR 23-MAY-1997; 97US-0047597.
PR 23-MAY-1997; 97US-0047598.
PR 23-MAY-1997; 97US-0047599.
PR 23-MAY-1997; 97US-0047600.
PR 23-MAY-1997; 97US-0047601.
PR 23-MAY-1997; 97US-0047612.
PR 23-MAY-1997; 97US-0047613.
PR 23-MAY-1997; 97US-0047614.
PR 23-MAY-1997; 97US-0047615.
PR 23-MAY-1997; 97US-0047617.
PR 23-MAY-1997; 97US-0047618.
PR 23-MAY-1997; 97US-0047632.
PR 23-MAY-1997; 97US-0047633.
PR 06-JUN-1997; 97US-0048964.
PR 06-JUN-1997; 97US-0048974.
PR 13-JUN-1997; 97US-0049610.
PR 08-JUL-1997; 97US-0051926.
PR 16-JUL-1997; 97US-0052874.
PR 18-AUG-1997; 97US-0055724.
PR 22-AUG-1997; 97US-0056630.
PR 22-AUG-1997; 97US-0056631.
PR 22-AUG-1997; 97US-0056632.
PR 22-AUG-1997; 97US-0056636.
PR 22-AUG-1997; 97US-0056637.
PR 22-AUG-1997; 97US-0056662.
PR 22-AUG-1997; 97US-0056664.
PR 22-AUG-1997; 97US-0056845.
PR 22-AUG-1997; 97US-0056862.
PR 22-AUG-1997; 97US-0056864.
PR 22-AUG-1997; 97US-0056872.
PR 22-AUG-1997; 97US-0056874.
PR 22-AUG-1997; 97US-0056875.
PR 22-AUG-1997; 97US-0056876.
PR 22-AUG-1997; 97US-0056877.
PR 22-AUG-1997; 97US-0056878.
PR 22-AUG-1997; 97US-0056879.
PR 22-AUG-1997; 97US-0056880.
PR 22-AUG-1997; 97US-0056881.
PR 22-AUG-1997; 97US-0056882.
PR 22-AUG-1997; 97US-0056884.
PR 22-AUG-1997; 97US-0056886.
PR 22-AUG-1997; 97US-0056887.
PR 22-AUG-1997; 97US-0056888.
PR 22-AUG-1997; 97US-0056889.
PR 22-AUG-1997; 97US-0056892.
PR 22-AUG-1997; 97US-0056893.
PR 22-AUG-1997; 97US-0056894.
PR 22-AUG-1997; 97US-0056903.
PR 22-AUG-1997; 97US-0056908.
PR 22-AUG-1997; 97US-0056909.
PR 22-AUG-1997; 97US-0056910.
PR 22-AUG-1997; 97US-0056911.
PR 05-SEP-1997; 97US-0057650.
PR 05-SEP-1997; 97US-0057669.
PR 05-SEP-1997; 97US-0057761.
PR 12-SEP-1997; 97US-0058785.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Bednarik DP, Brewer LA, Carter KC, Duan R, Ebner R, Endress GA;
PI Feng P, Ferrie AM, Fischer CL, Florence KA, Greene JM, Hu JS;
PI Kyaw H, Lafleur DW, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA;
PI Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;
XX
DR WPI; 1998-506364/43.
DR N-PSDB; AAV59765.
XX
PT New isolated human genes and the secreted polypeptide(s) they encode
PT - useful for diagnosis and treatment of e.g. cancers, neurological
PT disorders, immune diseases, inflammation or blood disorders
XX
PS Claim 1; Page 686-687; 721pp; English.
XX
CC This sequence represents a secreted human protein encoded by the nucleic
CC acid molecule designated Gene 106 from the human cDNA clone HT3AM65
CC (deposited as clone ATCC 97901 and ATCC 209047).
CC The gene can be used to generate fusion proteins by linking to the gene
CC to a human immunoglobulin Fc portion (e.g. AAV59502) for increasing the
CC stability of the fused protein as compared to the human protein only.
CC The invention relates to 186 novel genes and their fragments (nucleic
CC acid sequences: AAV59511-V59812; amino acid sequences AAW74731-W75026)
CC which are useful for preventing, treating or ameliorating medical
CC conditions e.g. by protein or gene therapy. Also, pathological
CC conditions can be diagnosed by determining the amount of the new
CC polypeptides in a sample or by determining the presence of mutations in
CC the new polynucleotides. Specific uses are described for each of the 186
CC polynucleotides, based on which tissues they are most highly expressed in
CC (see AAV59511 for described uses).
XX
SQ Sequence 263 AA;

Query Match 62.0%; Score 1344; DB 19; Length 263;
Best Local Similarity 100.0%; Pred. No. 7.7e-100;
Matches 262; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 150 MDCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNTVDLSSFDFR 209
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 MDCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNTVDLSSFDFR 60

```
Qy     210 TGKMMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKVKS 269
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TGKMMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKVKS 120

Qy     270 DPQRMNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSAVASAL 329
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 DPQRMNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSAVASAL 180

Qy     330 HTSSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEALMA 389
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 HTSSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEALMA 240

Qy     390 DILSRAADTEEMDIEMDSGDEA 411
           ||||||||||||||||||||||
Db     241 DILSRAADTEEMDIEMDSGDEA 262
```

```
RESULT 9
AAE22578
ID   AAE22578 standard; Protein; 282 AA.
XX
AC   AAE22578;
XX
DT   26-JUL-2002  (first entry)
XX
DE   Xenopus laevis MBD3 protein.
XX
KW   Gene expression; cellular chromatin; methyl CpG binding domain; cancer;
KW   localisation domain; diabetic retinopathy; ischaemia; HIV infection;
KW   human immuno deficiency virus; macular degeneration; vascular disease;
KW   rheumatoid arthritis; psoriasis; Alzheimer's disease; muscular dystrophy;
KW   sickle cell anaemia; stroke; neurodegenerative disease; cystic fibrosis;
KW   gene therapy; cytostatic; antidiabetic; ophthalmological; vasotropic;
KW   neuroprotective; nootropic; cerebroprotective; antibacterial; antifungal;
KW   antiviral; MBD3 protein.
XX
OS   Xenopus laevis.
XX
FH   Key             Location/Qualifiers
FT   Binding-site    1..69
FT                   /label= Methyl_CpG_binding_domain
FT   Region          7..9
FT                   /label= Beta1_helix
FT   Region          14..19
FT                   /label= Beta2_helix
FT   Region          20..30
FT                   /label= L1_loop
FT   Region          31..36
FT                   /label= Beta3_helix
FT   Region          40..42
FT                   /label= Beta4_helix
FT   Region          46..52
FT                   /label= Alpha1_helix
FT   Region          53..59
FT                   /label= L2_loop
FT   Region          61..68
FT                   /label= Hairpin_loop
XX
PN   WO200226960-A2.
XX
PD   04-APR-2002.
XX
PF   28-SEP-2001; 2001WO-US42377.
XX
PR   29-SEP-2000; 2000US-236884P.
XX
PA   (SANG-) SANGAMO BIOSCIENCES INC.
XX
PI   Wolffe AP,  Urnov F,  Lai A,  Raschke E;
XX
DR   WPI; 2002-372124/40.
XX
PT   Compartmentalizing a region of interest in cellular chromatin which
PT   facilitates the modulation of the expression of a gene comprises
PT   contacting the gene with a composition comprising a localization domain
PT   and a DNA binding domain -
XX
PS   Example 2; Fig 1B; 85pp; English.
XX
CC   The present invention relates to methods and compositions for regulating
CC   gene expression. In particular the method of compartmentalising a region
CC   of interest in cellular chromatin comprises contacting the region of
CC   interest with a composition that binds to a binding site in cellular
CC   chromatin, where the binding site is in a gene of interest and the
CC   composition comprises a localisation domain (e.g., methyl CpG binding
CC   domain obtained from MECP2, MBD1, MBD2, MBD3, dMBD-like and dMBD-like
CC   delta) and a DNA binding domain (or functional fragment). The method is
CC   useful for compartmentalising a region of interest in cellular chromatin
CC   which facilitates the modulation of the expression of a gene using a
CC   fusion molecule comprising a DNA binding domain and a localisation domain
CC   that binds to the chromatin. The fusion molecules or polypeptides can be
CC   used to prepare pharmaceutical compositions to prevent or treat cancer,
CC   ischaemia, diabetic retinopathy, macular degeneration, HIV infection,
CC   rheumatoid arthritis, psoriasis, sickle cell anaemia, vascular disease,
CC   Alzheimer's disease, muscular dystrophy, neurodegenerative diseases,
CC   cystic fibrosis, stroke, bacterial, viral or fungal infections. Sequences
CC   of the invention are also used in gene therapy. The present sequence is
CC   Xenopus laevis MBD3 LF protein. This sequence is used in the
CC   exemplification of the invention.
XX
SQ   Sequence   282 AA;
```

```
  Query Match             48.9%;  Score 1059;  DB 23;  Length 282;
  Best Local Similarity   75.4%;  Pred. No. 6.5e-77;
  Matches  202;  Conservative   33;  Mismatches   27;  Indels    6;  Gaps    2;

Qy     148 KRMDCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNTVDLSSFD 207
           || :| ||  ||||||| |:||||||||||||:||||||||||||||||||::|||:||
Db       4 KRWECSAL-QGWKKEEVTRRSGLSAGKSDVYYYSPSGKKFRSKPQLARYLGNSMDLSTFD 62

Qy     208 FRTGKMMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKV 267
           ||||||: ||: ||:||:| | |||:|||||||| ||:||||||||||||||||:|||
Db      63 FRTGKMLMSKINKNRQRMRYDGLNQSKGKPDLNTALPVRQTASIFKQPVTKVTNHPTNKV 122

Qy     268 KSDPQRMNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSAVAS 327
           |||||:  :||||||||:| ||:| |: |:::|||||||||||||  |||||||:||
Db     123 KSDPQKAVDQPRQLFWEKKLSGLNAFDIAEELVKTMELPKGLQGVGPGCTDETLLSAIAS 182

Qy     328 ALHTSSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEAL 387
           |||||: |||||:|||||||| |||||||||||||:|||||||||| ||||||||||
Db     183 ALHTSTMPITGQLSAAVEKNPGVWLNTSQPLCKAFMVTDEDIRKQEELVQQVRKKLEEAL 242

Qy     388 MADILSRAADTEE-----MDIEMDSGDE 410
           |||:|:   :  :     :| ::|  :|
Db     243 MADMLAHVEEISKDGGAPLDKDIDDEEE 270
```

```
RESULT 10
AAE22577
ID   AAE22577 standard; Protein; 303 AA.
XX
AC   AAE22577;
XX
DT   26-JUL-2002  (first entry)
XX
DE   Xenopus laevis MBD3 LF protein.
XX
KW   Gene expression; cellular chromatin; methyl CpG binding domain; cancer;
KW   localisation domain; diabetic retinopathy; ischaemia; HIV infection;
KW   human immuno deficiency virus; macular degeneration; vascular disease;
KW   rheumatoid arthritis; psoriasis; Alzheimer's disease; muscular dystrophy;
KW   sickle cell anaemia; stroke; neurodegenerative disease; cystic fibrosis;
KW   gene therapy; cytostatic; antidiabetic; ophthalmological; vasotropic;
KW   neuroprotective; nootropic; cerebroprotective; antibacterial; antifungal;
KW   antiviral; MBD3 protein.
```

```
XX
OS   Xenopus laevis.
XX
FH   Key             Location/Qualifiers
FT   Binding-site    1..90
FT                   /label= Methyl_CpG_binding_domain
FT   Region          7..9
FT                   /label= Beta1_helix
FT   Region          15..21
FT                   /label= Beta2_helix
FT   Region          22..31
FT                   /label= L1_loop
FT   Region          32..37
FT                   /label= Beta3_helix
FT   Region          61..63
FT                   /label= Beta4_helix
FT   Region          67..73
FT                   /label= Alpha1_helix
FT   Region          74..80
FT                   /label= L2_loop
FT   Region          82..89
FT                   /label= Hairpin_loop
XX
PN   WO200226960-A2.
XX
PD   04-APR-2002.
XX
PF   28-SEP-2001; 2001WO-US42377.
XX
PR   29-SEP-2000; 2000US-236884P.
XX
PA   (SANG-) SANGAMO BIOSCIENCES INC.
XX
PI   Wolffe AP,  Urnov F,  Lai A,  Raschke E;
XX
DR   WPI; 2002-372124/40.
XX
PT   Compartmentalizing a region of interest in cellular chromatin which
PT   facilitates the modulation of the expression of a gene comprises
PT   contacting the gene with a composition comprising a localization domain
PT   and a DNA binding domain -
XX
PS   Example 2; Fig 1B; 85pp; English.
XX
CC   The present invention relates to methods and compositions for regulating
CC   gene expression. In particular the method of compartmentalising a region
CC   of interest in cellular chromatin comprises contacting the region of
CC   interest with a composition that binds to a binding site in cellular
CC   chromatin, where the binding site is in a gene of interest and the
CC   composition comprises a localisation domain (e.g., methyl CpG binding
CC   domain obtained from MECP2, MBD1, MBD2, MBD3, dMBD-like and dMBD-like
CC   delta) and a DNA binding domain (or functional fragment). The method is
CC   useful for compartmentalising a region of interest in cellular chromatin
CC   which facilitates the modulation of the expression of a gene using a
CC   fusion molecule comprising a DNA binding domain and a localisation domain
CC   that binds to the chromatin. The fusion molecules or polypeptides can be
CC   used to prepare pharmaceutical compositions to prevent or treat cancer,
CC   ischaemia, diabetic retinopathy, macular degeneration, HIV infection,
CC   rheumatoid arthritis, psoriasis, sickle cell anaemia, vascular disease,
CC   Alzheimer's disease, muscular dystrophy, neurodegenerative diseases,
CC   cystic fibrosis, stroke, bacterial, viral or fungal infections. Sequences
CC   of the invention are also used in gene therapy. The present sequence is
CC   Xenopus laevis MBD3 LF protein. This sequence is used in the
CC   exemplification of the invention.
XX
SQ   Sequence   303 AA;

  Query Match             48.8%;  Score 1056.5;  DB 23;  Length 303;
  Best Local Similarity   70.5%;  Pred. No. 1.1e-76;
  Matches  203;  Conservative   33;  Mismatches   27;  Indels   25;  Gaps    2;

Qy    148 KRMDCPALPPGWKKEEVIRKSGLSAGKSDVYYFS----------------------PSGKKF 187
          ||  :| ||| ||||||| |:||||||||||||:|                      ||||||

Db      4 KRWECSALPQGWKKEEVTRRSGLSAGKSDVYYYSSSPSRYNRSLRDRVGCLNINPSGKKF 63

Qy    188 RSKPQLARYLGNTVDLSSFDFRTGKMMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQ 247
          ||||||||||||::|||:|||||||: ||: ||:|:| | |||:|||||||| ||:||
Db     64 RSKPQLARYLGNSMDLSTFDFRTGKMLMSKINKNRQRMRYDGLNQSKGKPDLNTALPVRQ 123

Qy    248 TASIFKQPVTKVTNHPSNKVKSDPQRMNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPK 307
          ||||||||||||||||:||||||||:  :|||||||||:| ||:| |: |:::|||||||
Db    124 TASIFKQPVTKVTNHPTNKVKSDPQKAVDQPRQLFWEKKLSGLNAFDIAEELVKTMELPK 183

Qy    308 GLQGVGPGSNDETLLSAVASALHTSSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDE 367
          |||||||  |||||||:|||||||: |||||:|||||||| ||||||||||||:||||
Db    184 GLQGVGPGCTDETLLSAIASALHTSTMPITGQLSAAVEKNPGVWLNTSQPLCKAFMVTDE 243

Qy    368 DIRKQEERVQQVRKKLEEALMADILSRAADTEE-----MDIEMDSGDE 410
          ||||||| ||||||||||||||:|:   :  :     :| ::|  :|
Db    244 DIRKQEELVQQVRKKLEEALMADMLAHVEEISKDGGAPLDKDIDDEEE 291


RESULT 11
AAY14198
ID   AAY14198 standard; Protein; 291 AA.
XX
AC   AAY14198;
XX
DT   28-JUL-1999  (first entry)
XX
DE   Human DNA demethylase, dMTase2, protien sequence.
XX
KW   DNA demethylase; dMTase1; dMTase2; human; cancer cell; beta-thalassemia;
KW   DNA methylation pattern; methylation inhibitor; gene therapy; diagnosis;
KW   genetic defect correction; sickle cell anaemia.
XX
OS   Homo sapiens.
XX
PN   WO9924583-A1.
XX
PD   20-MAY-1999.
XX
PF   12-NOV-1998;   98WO-CA01059.
XX
PR   11-MAY-1998;   98CA-2230991.
PR   12-NOV-1997;   97CA-2220805.
XX
PA   (UYMC-) UNIV MCGILL.
XX
PI   Bhattacharya S,  Ramchandani S,  Szyf M;
XX
DR   WPI; 1999-347283/29.
DR   N-PSDB; AAX61219.
XX
PT   Human and murine DNA demethylases useful in the diagnosis of cancer
XX
PS   Disclosure; Fig 9f; 114pp; English.
XX
CC   This sequence is the human DNA demethylase, designated dMTase2, of
CC   the invention. The DNA demethylase is overexpressed in cancer cells.
CC   Expression of the demethylase cDNA is useful to alter DNA methylation
CC   patterns of DNA in vitro in cells or in vivo in humans, animals and
CC   plants. The cDNA is in antisense orientation to inhibit demethylase in
CC   cancer cells for therapeutic purposes. The demethylase is used to alter
CC   the differentiation state and to generate stem cells for therapeutics,
CC   cells for animal cloning and to improve expression of foreign genes. The
CC   cDNA can also be used for recombinant production of large amounts of the
CC   demethylase. The protein can be used to raise antibodies against
CC   demethylase. It can also be used for high throughput screening of
CC   demethylase inhibitors, and for obtaining the x-ray crystal structure.
CC   The demethylase cDNA and protein are also useful for changing the state
CC   of differentiation of a cell to allow gene therapy, stem cell selection
CC   or cell cloning; or for inhibiting methylation in cancer cells using
CC   vector mediated gene therapy. Antagonists or inhibitors of the
CC   demethylase can be used to manufacture medicaments for cancer treatment,
CC   for restoring an aberrant methylation patterns or changing methylation
```

```
CC   patterns in patient DNA. Change of the methylation pattern activates a
CC   silent gene permitting the correction of a genetic defect, such as
CC   beta-thalassemia or sickle cell anaemia. The cDNA can be used as a
CC   template to design antisense oligonucleotides and ribozymes. The cDNA can
CC   also be used in two-hybrid systems in yeast to identify proteins
CC   interacting with demethylase. Determining the level of DNA methylase
CC   expression in a cell can be used as a method for diagnosis of cancer,
CC   where overexpression is indicative of cancer cells.
XX
SQ   Sequence   291 AA;

  Query Match             47.5%;  Score 1030;  DB 20;  Length 291;
  Best Local Similarity   72.0%;  Pred. No. 1.5e-74;
  Matches  198;  Conservative   34;  Mismatches   31;  Indels   12;  Gaps    2;

Qy     148 KRMDCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNTVDLSSFD 207
           || :||||| ||::||| |:|||||  ||:|:||||||||||||||||| ::|||:||
Db       4 KRWECPALPQGWEREEVPRRSGLSAGHRDVFYYSPSGKKFRSKPQLARYLGGSMDLSTFD 63

Qy     208 FRTGKMMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKV 267
           ||||||: ||: |::||:| |  || |||||||| ||:|||||||||||||:||||||||
Db      64 FRTGKMLMSKMNKSRQRVRYDSSNQVKGKPDLNTALPVRQTASIFKQPVTKITNHPSNKV 123

Qy     268 KSDPQRMNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSAVAS 327
           |||||:  :|||||||||:| ||:| |: |:::|||:|||||||||||  |||||||:||
Db     124 KSDPQKAVDQPRQLFWEKKLSGLNAFDIAEELVKTMDLPKGLQGVGPGCTDETLLSAIAS 183

Qy     328 ALHTSSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEAL 387
           |||||: |||||:|||||||| |||||:|||||||:|||||||||| |||||||:|||||
Db     184 ALHTSTMPITGQLSAAVEKNPGVWLNTTQPLCKAFMVTDEDIRKQEELVQQVRKRLEEAL 243

Qy     388 MADILSR----AADTE--------EMDIEMDSGDE 410
           |||:|:     | | |        | | | |  :|
Db     244 MADMLAHVEELARDGEAPLDKACAEDDDEEDEEEE 278


RESULT 12
AAY14200
ID   AAY14200 standard; Protein; 285 AA.
XX
AC   AAY14200;
XX
DT   28-JUL-1999  (first entry)
XX
DE   Mouse DNA demethylase, dMTase2, protien sequence.
XX
KW   DNA demethylase; dMTase1; dMTase2; mouse; cancer cell; beta-thalassemia;
KW   DNA methylation pattern; methylation inhibitor; gene therapy; diagnosis;
KW   genetic defect correction; sickle cell anaemia.
XX
OS   Mus sp.
XX
PN   WO9924583-A1.
XX
PD   20-MAY-1999.
XX
PF   12-NOV-1998;   98WO-CA01059.
XX
PR   11-MAY-1998;   98CA-2230991.
PR   12-NOV-1997;   97CA-2220805.
XX
PA   (UYMC-) UNIV MCGILL.
XX
PI   Bhattacharya S,  Ramchandani S,  Szyf M;
XX
DR   WPI; 1999-347283/29.
DR   N-PSDB; AAX61221.
XX
PT   Human and murine DNA demethylases useful in the diagnosis of cancer
XX
PS   Disclosure; Fig 9n; 114pp; English.
XX
CC   This sequence is the mouse DNA demethylase, designated dMTase2, of
CC   the invention. The DNA demethylase is overexpressed in cancer cells.
CC   Expression of the demethylase cDNA is useful to alter DNA methylation
CC   patterns of DNA in vitro in cells or in vivo in humans, animals and
CC   plants. The cDNA is in antisense orientation to inhibit demethylase in
CC   cancer cells for therapeutic purposes. The demethylase is used to alter
CC   the differentiation state and to generate stem cells for therapeutics,
CC   cells for animal cloning and to improve expression of foreign genes. The
CC   cDNA can also be used for recombinant production of large amounts of the
CC   demethylase. The protein can be used to raise antibodies against
CC   demethylase. It can also be used for high throughput screening of
CC   demethylase inhibitors, and for obtaining the x-ray crystal structure.
CC   The demethylase cDNA and protein are also useful for changing the state
CC   of differentiation of a cell to allow gene therapy, stem cell selection
CC   or cell cloning; or for inhibiting methylation in cancer cells using
CC   vector mediated gene therapy. Antagonists or inhibitors of the
CC   demethylase can be used to manufacture medicaments for cancer treatment,
CC   for restoring an aberrant methylation patterns or changing methylation
CC   patterns in patient DNA. Change of the methylation pattern activates a
CC   silent gene permitting the correction of a genetic defect, such as
CC   beta-thalassemia or sickle cell anaemia. The cDNA can be used as a
CC   template to design antisense oligonucleotides and ribozymes. The cDNA can
CC   also be used in two-hybrid systems in yeast to identify proteins
CC   interacting with demethylase. Determining the level of DNA methylase
CC   expression in a cell can be used as a method for diagnosis of cancer,
CC   where overexpression is indicative of cancer cells.
XX
SQ   Sequence   285 AA;

  Query Match             47.1%;  Score 1021.5;  DB 20;  Length 285;
  Best Local Similarity   75.2%;  Pred. No. 6.8e-74;
  Matches  191;  Conservative   35;  Mismatches   25;  Indels    3;  Gaps    1;

Qy     148 KRMDCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNTVDLSSFD 207
           || :||||| ||::||| |:|||||  ||:|:||||||||||||||||| ::|||:||
Db       4 KRWECPALPQGWEREEVPRRSGLSAGHRDVFYYSPSGKKFRSKPQLARYLGGSMDLSTFD 63

Qy     208 FRTGKMMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKV 267
           ||||||: :|: |::||:| |  || |||||||| ||:|||||||||||:|||||||||
Db      64 FRTGKMLMNKMNKSRQRVRYDSSNQVKGKPDLNTALPVRQTASIFKQPVTKITNHPSNKV 123

Qy     268 KSDPQRMNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSAVAS 327
           |||||:  :|||||||||:| |||| |: |::::||:|||||||||||  |||||||:||
Db     124 KSDPQKAVDQPRQLFWEKKLSGLSAFDIAEELVRTMDLPKGLQGVGPGCTDETLLSAIAS 183

Qy     328 ALHTSSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEAL 387
           |||||: |||||:|||||||| |||||:|||||||:|||:|||||||| ||||||:|||||
Db     184 ALHTSTLPITGQLSAAVEKNPGVWLNTAQPLCKAFMVTDDDIRKQEELVQQVRKRLEEAL 243

Qy     388 MADILSRAADTEEM 401
           |||:|   |  ||:
Db     244 MADML---AHVEEL 254


RESULT 13
AAY07107
ID   AAY07107 standard; Protein; 200 AA.
XX
AC   AAY07107;
XX
DT   02-JUL-1999  (first entry)
XX
DE   Colon cancer associated antigen precursor sequence.
XX
KW   Cancer associated antigen; diagnosis; research; treatment; human;
KW   breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW   prostate cancer.
XX
OS   Homo sapiens.
XX
PN   WO9904265-A2.
XX
PD   28-JAN-1999.
XX
```

```
PF   15-JUL-1998;   98WO-US14679.
XX
PR   22-JUN-1998;   98US-0102322.
PR   17-JUL-1997;   97US-0896164.
PR   10-OCT-1997;   97US-0061599.
PR   10-OCT-1997;   97US-0061765.
PR   10-OCT-1997;   97US-0948705.
PR   11-OCT-1997;   97GB-0021697.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
PI   Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;
PI   Tureci O;
XX
DR   WPI; 1999-132448/11.
XX
PT   New isolated cancer associated nucleic acids and polypeptides -
PT   isolated using sera from cancer patients, used to develop products
PT   for the diagnosis, monitoring or treatment of cancers
XX
PS   Disclosure; Page 677; 787pp; English.
XX
CC   The invention relates to a method for diagnosing a disorder characterised
CC   by expression of a human cancer associated antigen precursor coded for by
CC   a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC   biological sample isolated from a subject with an agent that specifically
CC   binds to the NAM, an expression product or a fragment of an expression
CC   product complexed with an HLA molecule; and (b) determining the
CC   interaction between the agent and the NAM or the expression product as a
CC   determination of the disorder. The products and methods can be used in
CC   the diagnosis, monitoring, research, or treatment of conditions
CC   characterised by the expression of various cancer associated antigens.
CC   The invention provides nucleic acid sequences and encoded polypeptides
CC   which are cancer associated antigen precursors expressed in human breast
CC   cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC   lung cancer.
XX
SQ   Sequence   200 AA;

  Query Match             46.7%;  Score 1013;  DB 20;  Length 200;
  Best Local Similarity   100.0%;  Pred. No. 2.1e-73;
  Matches  200;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     212 KMMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKVKSDP 271
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 KMMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKVKSDP 60

Qy     272 QRMNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSAVASALHT 331
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 QRMNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSAVASALHT 120

Qy     332 SSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEALMADI 391
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 SSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEALMADI 180

Qy     392 LSRAADTEEMDIEMDSGDEA 411
         ||||||||||||||||||||
Db     181 LSRAADTEEMDIEMDSGDEA 200


RESULT 14
AAB92997
ID   AAB92997 standard; Protein; 223 AA.
XX
AC   AAB92997;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human protein sequence SEQ ID NO:11731.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI   Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 11731; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence   223 AA;

  Query Match             34.7%;  Score 753;  DB 22;  Length 223;
  Best Local Similarity   70.5%;  Pred. No. 1.8e-52;
  Matches  148;  Conservative   25;  Mismatches   25;  Indels   12;  Gaps    2;

Qy     213 MMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKVKSDPQ 272
           |: ||: |::||:| |  || |||||||| ||:|||||||||||||:|||||||||||||
Db       1 MLMSKMNKSRQRVRYDSSNQVKGKPDLNTALPVRQTASIFKQPVTKITNHPSNKVKSDPQ 60

Qy     273 RMNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSAVASALHTS 332
           :  :|||||||||:| ||:| |: |:::|||:|||||||||||  |||||||:|||||||
Db      61 KAVDQPRQLFWEKKLSGLNAFDIAEELVKTMDLPKGLQGVGPGCTDETLLSAIASALHTS 120

Qy     333 SAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEALMADIL 392
           : |||||:|||||||| |||||:|||||||:|||||||||| ||||||:|||||||:|
Db     121 TMPITGQLSAAVEKNPGVWLNTTQPLCKAFMVTDEDIRKQEELVQQVRKRLEEALMADML 180

Qy     393 SR----AADTE--------EMDIEMDSGDE 410
           :     | | |         | | | |  :|
Db     181 AHVEELARDGEAPLDKACAEDDDEEDEEEE 210
```

```
RESULT 15
AAB93733
ID   AAB93733 standard; Protein; 218 AA.
XX
AC   AAB93733;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human protein sequence SEQ ID NO:13377.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 13377; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence   218 AA;

  Query Match             34.1%;  Score 739;  DB 22;  Length 218;
  Best Local Similarity   70.7%;  Pred. No. 2.3e-51;
  Matches  145;  Conservative   24;  Mismatches   24;  Indels   12;  Gaps    2;

Qy    218 LQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKVKSDPQRMNEQ 277
          : |::||:| |  || |||||||| ||:|||||||||||||:|||||||||||||:  :|
Db      1 MNKSRQRVRYDSSNQVKGKPDLNTALPVRQTASIFKQPVTKITNHPSNKVKSDPQKAVDQ 60

Qy    278 PRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSAVASALHTSSAPIT 337
          |||||||:| ||:| |: |:::|||:|||||||||||  |||||||:|||||||: |||
Db     61 PRQLFWEKKLSGLNAFDIAEELVKTMDLPKGLQGVGPGCTDETLLSAIASALHTSTMPIT 120

Qy    338 GQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEALMADILSR--- 394
          ||:|||||||| |||||:|||||||:|||||||||||| ||||||:|||||||:|:|:
Db    121 GQLSAAVEKNPGVWLNTTQPLCKAFMVTDEDIRKQEELVQQVRKRLEEALMADMLAHVEE 180

Qy    395 -AADTE--------EMDIEMDSGDE 410
           | | |         | | | |  :|
Db    181 LARDGEAPLDKACAEDDDEEDEEEE 205


Search completed: March 12, 2003, 05:40:34
Job time : 77.8484 secs
```

GenCore version 5.1.4_p5_4578

OM protein - protein search, using sw model

Run on: March 12, 2003, 03:11:54 ; Search time 10.5111 Seconds
(without alignments)
1033.836 Million cell updates/sec

Title: US-09-554-414B-2_COPY_150_411
Perfect score: 1344
Sequence: 1 MDCPALPPGWKKEEVIRKSG.....LSRAADTEEMDIEMDSGDEA 262

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	177	13.2	486	1	MEC2_HUMAN	P51608 homo sapien
2	167	12.4	484	1	MEC2_MOUSE	Q9z2d6 mus musculu
3	167	12.4	492	1	MEC2_RAT	Q00566 rattus norv
4	107.5	8.0	1039	1	MSL1_DROME	P50535 drosophila
5	97	7.2	448	1	YAP1_CHICK	P46936 gallus gall
6	95	7.1	612	1	ADF1_CANAL	P46589 candida alb
7	95	7.1	794	1	PMS1_SCHPO	P54280 schizosacch
8	95	7.1	1938	1	MYHD_HUMAN	Q9ukx3 homo sapien
9	93.5	7.0	1024	1	RIP3_MOUSE	P97434 mus musculu
10	92.5	6.9	459	1	ZPR1_MOUSE	Q62384 mus musculu
11	92.5	6.9	675	1	NED1_MOUSE	P33215 mus musculu
12	92	6.8	498	1	MEFA_MOUSE	Q60929 mus musculu
13	92	6.8	2774	1	MAPA_RAT	P34926 rattus norv
14	90.5	6.7	449	1	AROA_PSES2	P56952 pseudomonas
15	90	6.7	386	1	Y364_RICPR	Q9zdg5 rickettsia
16	89.5	6.7	823	1	NSP1_YEAST	P14907 saccharomyc
17	89.5	6.7	885	1	ASE1_YEAST	P50275 saccharomyc
18	89.5	6.7	2492	1	ATRX_HUMAN	P46100 homo sapien
19	89	6.6	507	1	MEFA_HUMAN	Q02078 homo sapien
20	89	6.6	1290	1	XCPC_XENLA	P50532 xenopus lae
21	89	6.6	2845	1	APC_MOUSE	Q61315 mus musculu
22	88.5	6.6	363	1	SP42_YEAST	P36094 saccharomyc
23	88.5	6.6	939	1	ST20_YEAST	Q03497 saccharomyc
24	88.5	6.6	1029	1	RIP3_RAT	Q9ere6 rattus norv
25	88	6.5	970	1	Y852_HUMAN	Q9y6x9 homo sapien
26	87.5	6.5	284	1	TPM2_DROME	P09491 drosophila
27	87.5	6.5	558	1	ORC2_XENLA	Q91628 xenopus lae
28	87.5	6.5	642	1	PHR_NEUCR	P27526 neurospora
29	87.5	6.5	1387	1	TROP_HUMAN	Q12816 homo sapien
30	87.5	6.5	1746	1	TENA_PIG	Q29116 sus scrofa
31	87	6.5	217	1	EVG1_HUMAN	Q9bze7 homo sapien
32	87	6.5	432	1	TIG_HAEIN	P44837 haemophilus
33	87	6.5	634	1	YKCA_CAEEL	P42083 caenorhabdi
34	87	6.5	861	1	ORC1_HUMAN	Q13415 homo sapien
35	86.5	6.4	407	1	YNH8_YEAST	P53939 saccharomyc
36	86.5	6.4	459	1	ZPR1_HUMAN	O75312 homo sapien
37	86.5	6.4	834	1	SRC1_YEAST	Q03707 saccharomyc
38	86.5	6.4	986	1	GM13_RAT	Q62839 rattus norv
39	86.5	6.4	1301	1	PTP9_DROME	P35832 drosophila
40	86.5	6.4	2095	1	RP1_MOUSE	P56716 mus musculu
41	86	6.4	310	1	FIRA_PLAFF	P06916 plasmodium
42	86	6.4	411	1	CPR1_PETCR	Q99089 petroselinu
43	86	6.4	559	1	PAXI_CHICK	P49024 gallus gall
44	86	6.4	576	1	UN87_CAEEL	P37806 caenorhabdi
45	86	6.4	622	1	YA41_HUMAN	Q9upw0 homo sapien

ALIGNMENTS

RESULT 1
MEC2_HUMAN
ID MEC2_HUMAN STANDARD; PRT; 486 AA.
AC P51608; O15233;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Methyl-CpG-binding protein 2 (MeCP-2 protein) (MeCP2).
GN MECP2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Kudo S., Fukuda M.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Thiesen J., Straetling W.H.;
RL Submitted (APR-1997) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=97130625; PubMed=8976388;
RA Vilain A., Apiou F., Vogt N., Dutrillaux B., Malfoy B.;
RT "Assignment of the gene for methyl-CpG-binding protein 2 (MECP2) to
RT human chromosome band Xq28 by in situ hybridization.";
RL Cytogenet. Cell Genet. 74:293-294(1996).
RN [4]
RP SEQUENCE FROM N.A.
RA Reichwald K., Rosenthal A., Kioschis P., Platzer M.;
RT "Mapping and sequence analysis of the human MeCP2 locus.";
RL Submitted (OCT-1997) to the EMBL/GenBank/DDBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE=99299240; PubMed=10369871;
RA Coy J.F., Sedlacek Z., Baechner D., Delius H., Poustka A.;
RT "A complex pattern of evolutionary conservation and alternative
RT polyadenylation within the long 3'-untranslated region of the
RT methyl-CpG-binding protein 2 gene (MeCP2) suggests a regulatory role
RT in gene expression.";
RL Hum. Mol. Genet. 8:1253-1262(1999).
RN [6]
RP SEQUENCE OF 10-486 FROM N.A.
RC TISSUE=Skeletal muscle;
RX MEDLINE=96327611; PubMed=8672133;
RA D'Esposito M., Quaderi N.A., Ciccodicola A., Bruni P., Esposito T.,
RA D'Urso M., Brown S.D.M.;
RT "Isolation, physical mapping, and Northern analysis of the X-linked
RT human gene encoding methyl CpG-binding protein, MECP2.";
RL Mamm. Genome 7:533-535(1996).
RN [7]
RP SEQUENCE OF 10-486 FROM N.A.
RA Reichwald K., Bauer D., Brenner V., Drescher B., Coy J.,
RA Kioschis P., Korn B., Nyakatura G., Platzer M., Poustka A.,

```
RA   Sandoval N., Rosenthal A.;
RL   Submitted (DEC-1996) to the EMBL/GenBank/DDBJ databases.
RN   [8]
RP   VARIANTS RTT TRP-106; CYS-133; SER-155 AND MET-158.
RX   MEDLINE=99438392; PubMed=10508514;
RA   Amir R.E., Van den Veyver I.B., Wan M., Tran C.Q., Francke U.,
RA   Zoghbi H.Y.;
RT   "Rett syndrome is caused by mutations in X-linked MECP2, encoding
RT   methyl-CpG-binding protein 2.";
RL   Nat. Genet. 23:185-188(1999).
RN   [9]
RP   VARIANT RTT VAL-140.
RX   MEDLINE=20465115; PubMed=11007980;
RA   Orrico A., Lam C., Galli L., Dotti M.T., Hayek G., Tong S.F.,
RA   Poon P.M., Zappella M., Federico A., Sorrentino V.;
RT   "MECP2 mutation in male patients with non-specific X-linked mental
RT   retardation.";
RL   FEBS Lett. 481:285-288(2000).
RN   [10]
RP   VARIANTS RTT W-106; F-124; C-13; C-134; R-152; M-158 AND C-306.
RX   MEDLINE=20439334; PubMed=10991688;
RA   Obata K., Matsuishi T., Yamashita Y., Fukuda T., Kuwajima K.,
RA   Horiuchi I., Nagamitsu S., Iwanaga R., Kimura A., Omori I., Endo S.,
RA   Mori K., Kondo I.;
RT   "Mutation analysis of the methyl-CpG binding protein 2 gene (MECP2) in
RT   patients with Rett syndrome.";
RL   J. Med. Genet. 37:608-610(2000).
RN   [11]
RP   VARIANTS RTT R-101; W-106; M-158 AND C-306, AND VARIANT K-397.
RX   MEDLINE=20439335; PubMed=10991689;
RA   Hampson K., Woods C.G., Latif F., Webb T.;
RT   "Mutations in the MECP2 gene in a cohort of girls with Rett
RT   syndrome.";
RL   J. Med. Genet. 37:610-612(2000).
RN   [12]
RP   VARIANT RTT VAL-140.
RX   MEDLINE=21664240; PubMed=11805248;
RA   Dotti M.T., Orrico A., De Stefano N., Battisti C., Sicurelli F.,
RA   Severi S., Lam C.W., Galli L., Sorrentino V., Federico A.;
RT   "A Rett syndrome MECP2 mutation that causes mental retardation in
RT   men.";
RL   Neurology 58:226-230(2002).
CC   -!- FUNCTION: CHROMOSOMAL PROTEIN THAT BINDS TO METHYLATED DNA. IT CAN
CC       BIND SPECIFICALLY TO A SINGLE METHYL-CPG PAIR. IT IS NOT
CC       INFLUENCED BY SEQUENCES FLANKING THE METHYL-CPGS. MEDIATES
CC       TRANSCRIPTIONAL REPRESSION THROUGH INTERACTION WITH HISTONE
CC       DEACETYLASE AND THE COREPRESSOR SIN3A.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR. COLOCALIZED WITH METHYL-CPG IN THE
CC       GENOME.
CC   -!- TISSUE SPECIFICITY: PRESENT IN ALL ADULT SOMATIC TISSUES TESTED.
CC   -!- DISEASE: DEFECTS IN MECP2 ARE THE CAUSE OF RETT SYNDROME (RTT), AN
CC       X-LINKED DOMINANT DISEASE. RTT IS A PROGRESSIVE NEUROLOGIC
CC       DEVELOPMENTAL DISORDER AND ONE OF THE MOST COMMON CAUSES OF MENTAL
CC       RETARDATION IN FEMALES. PATIENTS APPEAR TO DEVELOP NORMALLY UNTIL
CC       6 TO 18 MONTHS OF AGE, THEN GRADUALLY LOOSE SPEECH AND PURPOSEFUL
CC       HAND MOVEMENTS AND DEVELOP MICROCEPHALY, SEIZURES, AUTISM, ATAXIA,
CC       INTERMITTENT HYPERVENTILATION, AND STEREOTYPIC HAND MOVEMENTS.
CC       AFTER INITIAL REGRESSION, THE CONDITION STABILIZES AND PATIENTS
CC       USUALLY SURVIVE INTO ADULTHOOD.
CC   -!- SIMILARITY: CONTAINS 1 A.T HOOK DNA-BINDING REPEAT.
CC   -!- SIMILARITY: CONTAINS 1 METHYL-BINDING DOMAIN (MBD).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L37298; AAC32737.1; -.
DR   EMBL; Y12643; CAA73190.1; -.
DR   EMBL; X99686; CAA68001.1; -.
DR   EMBL; AF030876; AAC08757.1; -.
DR   EMBL; AF031078; AAC08758.1; -.
DR   EMBL; AJ132917; CAB46446.1; -.
DR   EMBL; X89430; CAA61599.1; -.
DR   EMBL; X94628; CAA64331.1; -.
DR   TRANSFAC; T04936; -.
DR   Genew; HGNC:6990; MECP2.
DR   MIM; 300005; -.
DR   MIM; 312750; -.
DR   InterPro; IPR000637; AT_hook.
DR   InterPro; IPR001739; Methyl-CpG_bind.
DR   Pfam; PF01429; MBD; 1.
DR   SMART; SM00384; AT_hook; 1.
DR   SMART; SM00391; MBD; 1.
KW   Transcription regulation; Repressor; DNA-binding; Nuclear protein;
KW   Disease mutation; Polymorphism.
FT   DOMAIN       96    149       MBD.
FT   DOMAIN      277    283       POLY-ALA.
FT   DOMAIN      366    372       POLY-HIS.
FT   DOMAIN      384    393       POLY-PRO.
FT   VARIANT     101    101       P -> R (IN RTT).
FT                                /FTId=VAR_010276.
FT   VARIANT     106    106       R -> W (IN RTT).
FT                                /FTId=VAR_010272.
FT   VARIANT     124    124       L -> F (IN RTT).
FT                                /FTId=VAR_010277.
FT   VARIANT     133    133       R -> C (IN RTT).
FT                                /FTId=VAR_010273.
FT   VARIANT     134    134       S -> C (IN RTT).
FT                                /FTId=VAR_010278.
FT   VARIANT     140    140       A -> V (IN RTT; AFFECTS WOMEN AND MEN).
FT                                /FTId=VAR_010279.
FT   VARIANT     152    152       P -> R (IN RTT).
FT                                /FTId=VAR_010280.
FT   VARIANT     155    155       F -> S (IN RTT).
FT                                /FTId=VAR_010274.
FT   VARIANT     158    158       T -> M (IN RTT).
FT                                /FTId=VAR_010275.
FT   VARIANT     201    201       A -> V.
FT                                /FTId=VAR_010281.
FT   VARIANT     306    306       R -> C (IN RTT).
FT                                /FTId=VAR_010282.
FT   VARIANT     397    397       E -> K.
FT                                /FTId=VAR_010283.
FT   CONFLICT     72     75       PAVP -> RLC (IN REF. 6).
FT   CONFLICT    290    290       E -> G (IN REF. 3).
SQ   SEQUENCE   486 AA;  52440 MW;  EB6A33233AEDA566 CRC64;

  Query Match             13.2%;  Score 177;  DB 1;  Length 486;
  Best Local Similarity   28.4%;  Pred. No. 6.7e-06;
  Matches   75;  Conservative   37;  Mismatches   86;  Indels   66;  Gaps   14;

Qy       2 DCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARY---LGNT-VDLSSF 57
           | | || || ::   |||| |||| |||  :| || |||| :|  |   :|:| :| : |
Db      96 DDPTLPEGWTRKLKQRKSGRSAGKYDVYLINPQGKAFRSKVELIAYFEKVGDTSLDPNDF 155

Qy      58 DFR-TGKMMPSKLQKNKQRLRNDP----LNQNKGKP-DLNTTLPIRQTASIFKQPVTKVT 111
           ||  ||:  ||: ::   :    |       : :|:|    || |   |:   :  | :|
Db     156 DFTVTGRGSPSRREQKPPKKPKSPKAPGTGRGRGRPKGSGTTRPKAATSEGVQ--VKRVL 213

Qy     112 NHPSNKVKSDPQRMNEQPRQLFWEKRLQGLSASDVTE-QIIK-----------TMELPKG 159
                 ||  : : : | |     : :|  |:  |:  :||            :||
Db     214 E------KSPGKLLVKMPFQTSPGGKAEGGGATTSTQVMVIKRPGRKRKAEADPQAIPK- 266

Qy     160 LQGVGPGSNDETLLSAVASALHTSSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDED 219
           :|  ||       | ||:|             ||  |  |              | :
Db     267 KRGRKPG-------SVVAAA-------------AAEAKKKA-------------VKESS 292

Qy     220 IRKQEERVQQVRK-KLEEALMADI 242
           ||  :| |  ::| |  | :  ::
Db     293 IRSVQETVLPIKKRKTRETVSIEV 316
```

```
RESULT 2
MEC2_MOUSE
ID   MEC2_MOUSE     STANDARD;      PRT;   484 AA.
AC   Q9Z2D6;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Methyl-CpG-binding protein 2 (MeCP-2 protein) (MeCP2).
GN   MECP2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6;
RX   MEDLINE=98449942; PubMed=9774669;
RA   Hendrich B., Bird A.;
RT   "Identification and characterization of a family of mammalian methyl-
RT   CpG binding proteins.";
RL   Mol. Cell. Biol. 18:6538-6547(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99299240; PubMed=10369871;
RA   Coy J.F., Sedlacek Z., Baechner D., Delius H., Poustka A.;
RT   "A complex pattern of evolutionary conservation and alternative
RT   polyadenylation within the long 3'-untranslated region of the methyl-
RT   CpG-binding protein 2 gene (MeCP2) suggests a regulatory role in gene
RT   expression.";
RL   Hum. Mol. Genet. 8:1253-1262(1999).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Reichwald K., Thiessen J., Wiehe T., Kioschis P., Straetling W.H.,
RA   Rosenthal A., Platzer M.;
RT   "Comparative analysis of the methyl CpG binding protein 2 locus in man
RT   and mouse reveals new untranslated sequences.";
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: CHROMOSOMAL PROTEIN THAT BINDS TO METHYLATED DNA. IT CAN
CC       BIND SPECIFICALLY TO A SINGLE METHYL-CPG PAIR. IT IS NOT
CC       INFLUENCED BY SEQUENCES FLANKING THE METHYL-CPGS. MEDIATES
CC       TRANSCRIPTIONAL REPRESSION THROUGH INTERACTION WITH HISTONE
CC       DEACETYLASE AND THE COREPRESSOR SIN3A (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: NUCLEAR. COLOCALIZED WITH METHYL-CPG IN THE
CC       GENOME.
CC   -!- SIMILARITY: CONTAINS 1 A.T HOOK DNA-BINDING REPEAT.
CC   -!- SIMILARITY: CONTAINS 1 METHYL-BINDING DOMAIN (MBD).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF072251; AAC68880.1; -.
DR   EMBL; AJ132922; CAB46495.1; -.
DR   EMBL; AF121351; AAF22116.1; -.
DR   EMBL; AF158181; AAF33024.1; -.
DR   MGD; MGI:99918; Mecp2.
DR   InterPro; IPR000637; AT_hook.
DR   InterPro; IPR001739; Methyl-CpG_bind.
DR   Pfam; PF01429; MBD; 1.
DR   PRINTS; PR00929; ATHOOK.
DR   SMART; SM00391; MBD; 1.
KW   Transcription regulation; Repressor; DNA-binding; Nuclear protein.
FT   DOMAIN       96    149       MBD.
FT   DOMAIN      277    283       POLY-ALA.
FT   DOMAIN      366    372       POLY-HIS.
FT   DOMAIN      384    391       POLY-PRO.
FT   DOMAIN      440    443       POLY-THR.
SQ   SEQUENCE   484 AA;  52307 MW;  62FD228F0118A49F CRC64;
```

```
  Query Match             12.4%;  Score 167;  DB 1;  Length 484;
  Best Local Similarity   26.3%;  Pred. No. 3.6e-05;
  Matches   71;  Conservative   31;  Mismatches   90;  Indels   78;  Gaps   12;

Qy       2 DCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARY---LGNT-VDLSSF 57
           | | || || ::   |||| |||| |||  :| || |||| :|  |   :|:| :| : |
Db      96 DDPTLPEGWTRKLKQRKSGRSAGKYDVYLINPQGKAFRSKVELIAYFEKVGDTSLDPNDF 155

Qy      58 DFR-TGKMMPSKLQKNKQ----------RLRNDPLNQNKGKPDLNTTLPIRQTASIFK 104
           ||  ||:  ||: ::              | |  |     |:|    :  ::    : |
Db     156 DFTVTGRGSPSRREQKPPKKPKSPKAPGTGRGRGRPKGSGTGRPKAAASEGVQVKRVLEK 215

Qy     105 QPVTKVTNHPSNKVKSDPQRMNEQPRQLFWEKRLQGLSASDVTEQIIK----------T 153
           |   |   |    :: |    |           | : :     :||
Db     216 SPGKLVVKMP---FQASPGGKGE----------GGGATTSAQVMVIKRPGRKRKAEADP 261

Qy     154 MELPKGLQGVGPGSNDETLLSAVASALHTSSAPITGQVSAAVEKNPAVWLNTSQPLCKAF 213
             :||  :|  ||       | ||:|             ||  |  |
Db     262 QAIPK-KRGRKPG------SVVAAA-----------AAEAKKKA------------ 287

Qy     214 IVTDEDIRKQEERVQQVRK-KLEEALMADI 242
           | :  ||   | |  ::| |  | :  ::
Db     288 -VKESSIRSVHETVLPIKKRKTRETVSIEV 316


RESULT 3
MEC2_RAT
ID   MEC2_RAT       STANDARD;      PRT;   492 AA.
AC   Q00566;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Methyl-CpG-binding protein 2 (MeCP-2 protein) (MeCP2).
GN   MECP2.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=Brain;
RX   MEDLINE=92298389; PubMed=1606614;
RA   Lewis J.D., Meehan R.R., Henzel W.J., Maurer-Fogy I., Jeppesen P.,
RA   Klein F., Bird A.;
RT   "Purification, sequence, and cellular localization of a novel
RT   chromosomal protein that binds to methylated DNA.";
RL   Cell 69:905-914(1992).
CC   -!- FUNCTION: CHROMOSOMAL PROTEIN THAT BINDS TO METHYLATED DNA. IT CAN
CC       BIND SPECIFICALLY TO A SINGLE METHYL-CPG PAIR. IT IS NOT
CC       INFLUENCED BY SEQUENCES FLANKING THE METHYL-CPGS. MEDIATES
CC       TRANSCRIPTIONAL REPRESSION THROUGH INTERACTION WITH HISTONE
CC       DEACETYLASE AND THE COREPRESSOR SIN3A (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: NUCLEAR. COLOCALIZED WITH METHYL-CPG IN THE
CC       GENOME.
CC   -!- TISSUE SPECIFICITY: PRESENT IN ALL ADULT SOMATIC TISSUES TESTED.
CC   -!- SIMILARITY: CONTAINS 1 A.T HOOK DNA-BINDING REPEAT.
CC   -!- SIMILARITY: CONTAINS 1 METHYL-BINDING DOMAIN (MBD).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M94064; AAA41584.1; -.
DR   PIR; A41907; A41907.
DR   InterPro; IPR001739; Methyl-CpG_bind.
DR   Pfam; PF01429; MBD; 1.
DR   SMART; SM00391; MBD; 1.
KW   Transcription regulation; Repressor; DNA-binding; Nuclear protein.
FT   DOMAIN       96    149       MBD.
```

```
FT   DOMAIN       366    372       POLY-HIS.
FT   DOMAIN       384    391       POLY-PRO.
FT   DOMAIN       443    451       POLY-THR.
SQ   SEQUENCE   492 AA;  53047 MW;  A67E705C68BA2D38 CRC64;

  Query Match             12.4%;  Score 167;  DB 1;  Length 492;
  Best Local Similarity   28.1%;  Pred. No. 3.7e-05;
  Matches   74;  Conservative   37;  Mismatches   88;  Indels   64;  Gaps   13;

Qy       2 DCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARY---LGNT-VDLSSF 57
           | | || || ::   |||| |||| |||  :| || |||| :|  |   :|:| :| : |
Db      96 DDPTLPEGWTRKLKQRKSGRSAGKYDVYLINPQGKAFRSKVELIAYFEKVGDTSLDPNDF 155

Qy      58 DFR-TGKMMPSKLQKNKQRLRNDP----LNQNKGKPDLNTTLPIRQTASIFKQPVTKVTN 112
           ||  ||:  ||: ::   :    |      : :|:|  : |   :  ||   | | :|
Db     156 DFTVTGRGSPSRREQKPPKKPKSPKAPGTGRGRGRPKGSGTGRPKAAASEGVQ-VKRVLE 214

Qy     113 HPSNKVKSDPQRMNEQPRQLFWEKRLQGLSASDVTE-QIIK-----------TMELPKGL 160
               ||  : : : | |     : :|  |:   :  :||             :||
Db     215 ------KSPGKLLVKMPFQASPGGKGEGGGATTSAQVMVIKRPGRKRKAEADPQAIPK-K 267

Qy     161 QGVGPGSNDETLLSAVASALHTSSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDI 220
           :|  ||       | ||:|             ||  |  |              | :  |
Db     268 RGRKPG-------SVVAAA------------AAEAKKKA-------------VKESSI 293

Qy     221 RKQEERVQQVRK-KLEEALMADI 242
           |  :| |  ::| |  | :  ::
Db     294 RSVQETVLPIKKRKTRETVSIEV 316


RESULT 4
MSL1_DROME
ID   MSL1_DROME     STANDARD;      PRT;  1039 AA.
AC   P50535;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   Male-specific lethal-1 protein.
GN   MSL-1.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC   Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC   Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE OF 85-1039 FROM N.A.
RC   STRAIN=Canton-S;
RX   MEDLINE=93314941; PubMed=8325488;
RA   Palmer M.J., Mergner V.A., Richman R., Manning J.E., Kuroda M.I.,
RA   Lucchesi J.C.;
RT   "The male-specific lethal-one (msl-1) gene of Drosophila melanogaster
RT   encodes a novel protein that associates with the X chromosome in
RT   males.";
RL   Genetics 134:545-557(1993).
RN   [2]
RP   REVISIONS, SEQUENCE FROM N.A.
RX   MEDLINE=95300219; PubMed=7781064;
RA   Kelley R.L., Solovyeva I., Lyman L.M., Richman R., Solovyev V.,
RA   Kuroda M.I.;
RT   "Expression of msl-2 causes assembly of dosage compensation
RT   regulators on the X chromosomes and female lethality in Drosophila.";
RL   Cell 81:867-877(1995).
CC   -!- FUNCTION: THE MSL PROTEINS ARE ESSENTIAL FOR ELEVATING
CC       TRANSCRIPTION OF THE SINGLE X CHROMOSOME IN THE MALE (X CHROMOSOME
CC       DOSAGE COMPENSATION). MSL-1 IS A PIONEER PROTEIN. MLE, MSL-1 AND
CC       MSL-3 ARE CO-LOCALIZED ON THE X CHROMOSOME. EACH OF THE MSL
CC       PROTEINS REQUIRES ALL THE OTHER MSLS FOR WILD-TYPE X-CHROMOSOME
CC       BINDING.
CC   -!- SUBUNIT: MSL-1 SEEMS TO FORM A TIGHT COMPLEX WITH MSL-2.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR; MSL-1 IS ASSOCIATED WITH HUNDREDS
CC       OF DISCRETE SITES ALONG THE LENGTH OF THE X CHROMOSOME IN MALES
CC       AND NOT IN FEMALES, AND IS ALSO ASSOCIATED WITH 10-20 AUTOSOMAL
CC       SITES IN MALES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L42514; AAA98918.1; -.
DR   FlyBase; FBgn0005617; msl-1.
KW   Nuclear protein.
FT   CONFLICT     188    193       PLPPAA -> HCHLLP (IN REF. 1).
FT   CONFLICT     492    492       L -> S (IN REF. 1).
FT   CONFLICT     670    670       I -> M (IN REF. 1).
SQ   SEQUENCE   1039 AA;  117412 MW;  4759E9B5EF6E9F14 CRC64;

  Query Match             8.0%;  Score 107.5;  DB 1;  Length 1039;
  Best Local Similarity   23.6%;  Pred. No. 2.4;
  Matches   48;  Conservative   26;  Mismatches   58;  Indels   71;  Gaps    7;

Qy      73 KQRLRNDPLNQNKGKP-----------DLNTTLPIRQTASIFKQPVTKV-TNHPSNKVKS 120
           |: |:  | :  |  |             :|||   || |   |  ||  || | | ::
Db     558 KETLKRQPEDAPKHLPKAVAPKVTKTSSRESTLPKANTADIKDAPAQKVIANHQSTKTQT 617

Qy     121 DP---QRMNEQPRQ--LFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDET---- 171
           ||   ||:  : || :  : |    : ||: :|           : | | |  ||
Db     618 DPVKTQRLQVKIRQYEMHPDMRTGSSAPSDIRKQ-----------KNVDPVSTPETKTIK 666

Qy     172 ----LLSAVASALHTSSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERV 227
               |::   :   || :|                                    || |
Db     667 SKSILVNDKKTTSETSQSP-----------------------------------DQEIDV 691

Qy     228 QQVRKKLEEALMADILSRAADTE 250
           : ||:|| | |  ::||::  ::
Db     692 ETVRRKLAEHLKKELLSQSHSSQ 714


RESULT 5
YAP1_CHICK
ID   YAP1_CHICK     STANDARD;      PRT;   448 AA.
AC   P46936;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   65 kDa Yes-associated protein (YAP65).
GN   YAP1 OR YAP65.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=White leghorn;
RX   MEDLINE=94309887; PubMed=8035999;
RA   Sudol M.;
RT   "Yes-associated protein (YAP65) is a proline-rich phosphoprotein that
RT   binds to the SH3 domain of the Yes proto-oncogene product.";
RL   Oncogene 9:2145-2152(1994).
CC   -!- FUNCTION: BINDS TO THE SH3 DOMAIN OF THE YES KINASE. ALSO BINDS TO
CC       OTHER SIGNALING MOLECULES THAT CONTAIN SH3 DOMAINS INCLUDING NCK,
CC       CRK AND SRC.
CC   -!- PTM: PHOSPHORYLATED ON SERINE RESIDUES.
CC   -!- SIMILARITY: CONTAINS 1 OR 2 WW DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
```

```
CC   ----------------------------------------------------------------------
DR   EMBL; X76483; CAA54021.1; -.
DR   InterPro; IPR005153; MbtH.
DR   InterPro; IPR002349; WW.
DR   InterPro; IPR001202; WW_Rsp5_WWP.
DR   Pfam; PF00397; WW; 1.
DR   Pfam; PF03621; MbtH; 1.
DR   PRINTS; PR00403; WWDOMAIN.
DR   SMART; SM00456; WW; 1.
DR   PROSITE; PS01159; WW_DOMAIN_1; 1.
DR   PROSITE; PS50020; WW_DOMAIN_2; 1.
KW   Phosphorylation.
FT   DOMAIN      169    202       WW.
SQ   SEQUENCE   448 AA;  47822 MW;  719CC8D0F879A38D CRC64;

  Query Match             7.2%;  Score 97;  DB 1;  Length 448;
  Best Local Similarity   22.0%;  Pred. No. 4.8;
  Matches   64;  Conservative   35;  Mismatches  114;  Indels   78;  Gaps   10;

Qy       2 DCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNTVDLSSFDFRT 61
           | | |||||:  :                :|||:         ||  | :| ::
Db     168 DVP-LPPGWEMAK----------------TPSGQ---------RYFLNHIDQTTTWQDP 200

Qy      62 GKMMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKV--- 118
           | | |  | |     : |: ||           | |:| : |||     |   |
Db     201 RKAMLS--QMNVTAPTSPPVQQNLMNSASAMNQRISQSAPV-KQPPPLAPQSPQGGVMGG 257

Qy     119 KSDPQRMNEQPRQLFWEKRLQGLSASDVTEQ------IIKTMELPKGLQG--VGPG---- 166
           |  |:   : :||  ||    |   ::  |       : |||   | |     ||
Db     258 SSSNQQQQMRLQQLQMEKERLRLKHQELLRQELALRSQLPTMEQDGGSQNPVSSPGMSQE 317

Qy     167 -------SNDETLLSAVASALHTSSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDED 219
                 |:|  | |    :   |:       | :| : |  :||:   :     ::  :
Db     318 LRTMTTNSSDPFLNSGTYHSRDESTDSGLSMSSYSVPRTPDDFLNSVDEMDTGDSISQSN 377

Qy     220 IRKQEERV-------------------------QQVRKKLEEALMADILS 244
           |   : |                         :::   |:||| :|||:
Db     378 IPSHQNRFPDYLEAIPGTNVDLGTLEGDGMNIEGEELMPSLQEALSSDILN 428


RESULT 6
ADF1_CANAL
ID   ADF1_CANAL     STANDARD;      PRT;   612 AA.
AC   P46589;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   Adherence factor (Adhesion and aggregation mediating surface antigen).
GN   ADF1 OR AAF1.
OS   Candida albicans (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5476;
RN   [1]
RP   SEQUENCE OF 35-612 FROM N.A.
RC   STRAIN=ATCC 36082;
RA   Edwards J.E.;
RL   Submitted (DEC-1994) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Jiang W., Finkler A., Koltin Y.;
RL   Submitted (JAN-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: SURFACE ANTIGEN MEDIATING ADHESION AND AGGREGATION IN
CC       S.CEREVISIAE.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U18983; AAA62506.1; -.
DR   EMBL; U44747; AAA86758.1; -.
KW   Cell adhesion.
FT   DOMAIN       64     69       POLY-GLN.
FT   DOMAIN      100    103       POLY-SER.
FT   DOMAIN      117    134       POLY-GLN.
FT   DOMAIN      269    272       POLY-ASN.
FT   DOMAIN      517    520       POLY-PRO.
FT   DOMAIN      579    582       POLY-PRO.
FT   DOMAIN      596    605       POLY-GLN.
FT   CONFLICT    142    142       F -> N (IN REF. 1).
FT   CONFLICT    451    451       T -> I (IN REF. 1).
SQ   SEQUENCE   612 AA;  68794 MW;  77B60E45C35C1B23 CRC64;

  Query Match             7.1%;  Score 95;  DB 1;  Length 612;
  Best Local Similarity   22.4%;  Pred. No. 10;
  Matches   51;  Conservative   38;  Mismatches   91;  Indels   48;  Gaps   10;

Qy      36 KKFRSKPQLARYLGNTVDLSSFDFRTGKMMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLP 95
           :: : :|   ::  ||:   ::      : ||: |   |   |        :| | :  |
Db     131 QQMQPQPNNMQFFDNTI--PNYLIMNQTISPSQTQTTAQ--PNISYYNYSTQPQLQSAQP 186

Qy      96 IRQTASIFKQPVTKVTNHPSNKVKSDPQRMNEQPRQLFWEKRLQGLSASDVTEQIIKTME 155
           |  :     ||  : |   ||: :|  |     :||     ::: |   |   :   |
Db     187 ISHSQ---PQPQPQATQPRSNRSRS--QTSFSKPR---GSRQVSGSGRSTGAK---KQSA 235

Qy     156 LPKGLQGVGPGSNDETLLSAVASALHTSSAPITG-----QVSAAVEKNPAVWLNTSQPLC 210
           :  | | ||  | :| :::||::  |::   |       ||| |     :: |
Db     236 ITSGSTGTGPARNADTGMTSVANSTSTTTMTTTNNNNKLSVSAPVN---VIYANL----- 287

Qy     211 KAFIVTDEDIRKQEERVQQVRKKLEEALMADILSRAADTEEMDIEMDS 258
                    ||:|||       | |  ||||    :: : : |
Db     288 -------------PERLQQV-------LPAPPLSRAPVRPDVTVNLTS 315


RESULT 7
PMS1_SCHPO
ID   PMS1_SCHPO     STANDARD;      PRT;   794 AA.
AC   P54280;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   DNA mismatch repair protein pms1.
GN   PMS1 OR SPAC19G12.02C.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=97403304; PubMed=9258673;
RA   Schar P., Baur M., Schneider C., Kohli J.;
RT   "Mismatch repair in Schizosaccharomyces pombe requires the mutL
RT   homologous gene pms1: molecular cloning and functional analysis.";
RL   Genetics 146:1275-1286(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=21848401; PubMed=11859360;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
```

```
RA   Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA   Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA   Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT   "The genome sequence of Schizosaccharomyces pombe.";
RL   Nature 415:871-880(2002).
CC   -!- FUNCTION: THIS PROTEIN IS INVOLVED IN THE REPAIR OF MISMATCHES
CC       IN DNA.
CC   -!- SIMILARITY: BELONGS TO THE DNA MISMATCH REPAIR MUTL/HEXB FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X96581; CAA65400.1; -.
DR   EMBL; Z97209; CAB10113.1; -.
DR   HSSP; P23367; 1BKN.
DR   InterPro; IPR003594; ATPbind_ATPase.
DR   InterPro; IPR002099; DNA_mis_repair.
DR   Pfam; PF01119; DNA_mis_repair; 1.
DR   Pfam; PF02518; HATPase_c; 1.
DR   TIGRFAMs; TIGR00585; mutl; 1.
DR   PROSITE; PS00058; DNA_MISMATCH_REPAIR_1; 1.
KW   DNA repair.
SQ   SEQUENCE   794 AA;  88009 MW;  A5D46FFFA077D8DC CRC64;

  Query Match             7.1%;  Score 95;  DB 1;  Length 794;
  Best Local Similarity   23.7%;  Pred. No. 14;
  Matches   46;  Conservative   32;  Mismatches   62;  Indels   54;  Gaps    9;

Qy      18 KSGLSAGKSDVYYFSPSGKKF---RSKPQLARYLGNTVDLSSFDFRTGKMMPSKLQKNKQ 74
           :|  |   |  |  ||  :|     ::| ::  :     |:    :  | :| :|||:
Db     373 ESETSDDSSFSYKRSPCKRKLVEATAQPAISTSVAEGASLA---QVSKPLPERLQKDSM 428

Qy      75 RLRNDPLNQN--------KGKPDL---NTTLPIRQTASI---FKQPVTKVTNHPSNKVKS 120
           | |: |||:         | |  |   :|   :::|       |||: | :::|:| : :
Db     429 R-RSSPLNEKVTASSERMKKKLALFASSTDTSMQKTIDSSFPLKQPINKPSSNPNNLLLN 487

Qy     121 DPQRMNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSAVASAL 180
           ||                    |  :  : ||: | : :: |    :  ||
Db     488 DP--------------------SPASTPVAKTINLNE-IESVHNAESVSTL-------- 517

Qy     181 HTSSAPITGQVSAA 194
             || | | | | |
Db     518 --SSIPRTEQTSVA 529


RESULT 8
MYHD_HUMAN
ID   MYHD_HUMAN     STANDARD;      PRT;  1938 AA.
AC   Q9UKX3; O95252;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Myosin heavy chain, skeletal muscle, extraocular (MyHC-eo).
GN   MYH13.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Extraocular muscle;
RX   MEDLINE=99318869; PubMed=10388558;
RA   Weiss A., Schiaffino S., Leinwand L.A.;
RT   "Comparative sequence analysis of the complete human sarcomeric myosin
RT   heavy chain family: implications for functional diversity.";
RL   J. Mol. Biol. 290:61-75(1999).
RN   [2]
RP   SEQUENCE OF 1917-1938 FROM N.A.
RC   TISSUE=Extraocular muscle;
RX   MEDLINE=99026150; PubMed=9806854;
RA   Winters L.M., Briggs M.M., Schachat F.;
RT   "The human extraocular muscle myosin heavy chain gene (MYH13) maps to
RT   the cluster of fast and developmental myosin genes on chromosome 17.";
RL   Genomics 54:188-189(1998).
CC   -!- FUNCTION: MUSCLE CONTRACTION.
CC   -!- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
CC       HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
CC       AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC   -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC   -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC       CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC       CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC   -!- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
CC       ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
CC   -!- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
CC       MEROMYOSIN (LMM) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN LATER BE
CC       SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
CC       SUBFRAGMENT (S2).
CC   -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF111782; AAD29948.1; -.
DR   EMBL; AF075248; AAC83241.1; -.
DR   HSSP; P13538; 2MYS.
DR   Genew; HGNC:7571; MYH13.
DR   MIM; 603487; -.
DR   InterPro; IPR000048; IQ_region.
DR   InterPro; IPR004009; Myosin_N.
DR   InterPro; IPR002928; Myosin_tail.
DR   InterPro; IPR001609; myosin_head.
DR   Pfam; PF00063; myosin_head; 1.
DR   Pfam; PF00612; IQ; 2.
DR   Pfam; PF01576; Myosin_tail; 1.
DR   Pfam; PF02736; Myosin_N; 1.
DR   PRINTS; PR00193; MYOSINHEAVY.
DR   ProDom; PD000355; myosin_head; 1.
DR   SMART; SM00015; IQ; 1.
DR   SMART; SM00242; MYSc; 1.
DR   PROSITE; PS50096; IQ; 1.
KW   Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW   Calmodulin-binding; ATP-binding; Methylation; Alkylation;
KW   Multigene family.
FT   DOMAIN        1    784       MYOSIN HEAD-LIKE.
FT   DOMAIN      785    814       IQ.
FT   DOMAIN      843   1938       COILED COIL (POTENTIAL).
FT   NP_BIND     179    186       ATP (POTENTIAL).
FT   DOMAIN      659    681       ACTIN-BINDING (BY SIMILARITY).
FT   DOMAIN      761    775       ACTIN-BINDING (BY SIMILARITY).
FT   MOD_RES     130    130       METHYLATION (TRI-) (POTENTIAL).
FT   MOD_RES     699    699       ALKYLATION (SH-1) (POTENTIAL).
FT   MOD_RES     709    709       ALKYLATION (SH-2) (POTENTIAL).
SQ   SEQUENCE   1938 AA;  223678 MW;  1F6D006416381CD5 CRC64;

  Query Match             7.1%;  Score 95;  DB 1;  Length 1938;
```

```
  Best Local Similarity   20.5%;  Pred. No. 44;
  Matches   61;  Conservative   49;  Mismatches  139;  Indels   48;  Gaps    7;

Qy       2 DCPALPPGWKKEEVIR---KSGLSAGKSDV-----YYFSPSGKKFRSKPQLARYLGNTVD 53
           ||  |   :::|:  :   :  ||   |:|      | : : ::     :  : |   :
Db    1343 DCDLLREQYEEEQEAKAELQRALSKANSEVAQWKTKYETDAIQRTEELEEAKKKLAQRLQ 1402

Qy      54 LSSFDFRTGKMMPSKLQKNKQRLRND--------------------PLNQNKGKPDLNT 92
            :     |     : |:| ||||: :                      :|   :
Db    1403 EAEEKTETANSKCASLEKTKQRLQGEVEDLMRDLERSHTACATLDKKQRNFDKVLAEWKQ 1462

Qy      93 TLPIRQTASIFKQPVTKVTNHPSNKVKSDPQRMNEQPRQLFWEKRLQGLSASDVTEQIIK 152
            |   |     |  ::  :    |:::  : : :|   |  | :      ||:|||| :
Db    1463 KLDESQAELEAAQKESRSLSTELFKMRNAYEEVVDQLETLRRENKNLQEEISDLTEQIAE 1522

Qy     153 TMELPKGLQGVGPGSNDETLLSAVASALHTSSAPITGQVSAAVEKNPAVWLNTSQ---PL 209
           |    | ||          : |:    | |  :   : | :      |   | |  ||    |
Db    1523 T---GKNLQ---EAEKTKKLVEQEKSDLQVALEEVEGSLEHEESKILRVQLELSQVKSEL 1576

Qy     210 CKAFIVTDEDIRKQEERVQQVRKKLEEALMADILSR----------AADTEEMDIEM 256
           :  |  ||:| : :   |:  : |:  | |:| ||            |  ||:|::
Db    1577 DRKVIEKDEEIEQLKRNSQRAAEALQSVLDAEIRSRNDALRLKKKMEGDLNEMEIQL 1633


RESULT 9
RIP3_MOUSE
ID   RIP3_MOUSE     STANDARD;      PRT;  1024 AA.
AC   P97434;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Rho-interacting protein 3 (p116RIP) (RIP3).
GN   RHOIP3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=97344280; PubMed=9199174;
RA   Gebbink M.F.B.G., Kranenburg O., Poland M., van Horck F.P.G.,
RA   Houssa B., Moolenaar W.H.;
RT   "Identification of a novel, putative Rho-specific GDP/GTP exchange
RT   factor and a RhoA-binding protein: control of neuronal morphology.";
RL   J. Cell Biol. 137:1603-1613(1997).
CC   -!- FUNCTION: RHOA-BINDING PROTEIN INVOLVED IN CONTROL OF THE ACTIN
CC       CYTOSKELETON. OVEREXPRESSION PROMOTES NEURONAL CELL FLATTENING AND
CC       NEURITE OUTGROWTH, PROBABLY BY COUNTERACTING RHOA-MEDIATED
CC       SIGNALING.
CC   -!- TISSUE SPECIFICITY: HIGHLY ENRICHED IN THE BRAIN.
CC   -!- SIMILARITY: CONTAINS 2 PH DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U73200; AAB18198.1; -.
DR   MGD; MGI:1349438; Rhoip3.
DR   InterPro; IPR001849; PH.
DR   Pfam; PF00169; PH; 2.
DR   SMART; SM00233; PH; 2.
DR   PROSITE; PS50003; PH_DOMAIN; 2.
KW   Guanine-nucleotide releasing factor; Repeat; Coiled coil.
FT   DOMAIN       43    150       PH 1.
FT   DOMAIN      386    482       PH 2.
FT   DOMAIN      164    172       PRO-RICH.
FT   DOMAIN      185    190       POLY-SER.
FT   DOMAIN      285    296       PRO-RICH.
FT   DOMAIN      671    976       COILED COIL (POTENTIAL).
SQ   SEQUENCE   1024 AA;  116363 MW;  A91BDBB955E45E02 CRC64;

  Query Match             7.0%;  Score 93.5;  DB 1;  Length 1024;
  Best Local Similarity   19.5%;  Pred. No. 25;
  Matches   61;  Conservative   53;  Mismatches  132;  Indels   67;  Gaps   11;

Qy       5 ALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNTVDLSSFDFR---- 60
           :|| | |           |  | :     |  :: :|: :  |  | :       :||
Db     492 SLPEGKNKSTSFETCSRSTEKQEAEPGEPDPEQKKSRARERRREGRSKTFDWAEFRPIQQ 551

Qy      61 --------------TG-----KMMPSKLQKNKQRLRNDPLNQ--------NKGKPDLNTT 93
                         :|     :  | :|:: : | | :|  :          |  |
Db     552 ALAQERASAVGSSDSGDPGCLEAEPGELERERARRREEPRKRFGMLDTIDGPGMEDTALR 611

Qy      94 LPIRQTASIFKQPVTKVTN-HPS-----NKVKSDPQRMNEQ----PRQLFWEKRLQGLSA 143
           : | ::  :    |  |  | |          ::|:: | |  :|    |  |  | | : ||
Db     612 MDIDRSPGLLGTPDLKTQNVHVEIEQRWHQVETTPLREEKQVPIAPLHLSLEDRSERLST 671

Qy     144 SDVTEQIIKTMELPKGLQGVGPGSN----DETLLSAVASALHTSSAPITGQV-SAAVEKN 198
            ::|  : | :|     |     |:    :  |   :  ||    :   | | |  |:
Db     672 HELTSLLEKELE-----QSQKEASDLLEQNRLLQDQLRVALGREQSAREGYVLQATCERG 726

Qy     199 PAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEALMADILS---------RAADT 249
           |    | |          ||:::| :|  :  :: :: |:|:  :          : |
Db     727 FAAMEETHQKKI-------EDLQRQHQRELEKLREEKDRLLAEETAATISAIEAMKNAHR 779

Qy     250 EEMDIEMDSGDEA 262
           |||: |::    :
Db     780 EEMERELEKSQRS 792


RESULT 10
ZPR1_MOUSE
ID   ZPR1_MOUSE     STANDARD;      PRT;   459 AA.
AC   Q62384;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Zinc-finger protein ZPR1 (Zinc finger protein 259).
GN   ZNF259 OR ZFP259 OR ZPR1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Galcheva-Gargova Z., Konstantinov K.N., Wu I.-H., Klier F.G.,
RA   Barrett T., Davis R.J.;
RT   "Binding of zinc finger protein ZPR1 to the epidermal growth
RT   factor receptor.";
RL   Science 272:1797-1802(1996).
RN   [2]
RP   SUBCELLULAR LOCATION.
RX   MEDLINE=98437195; PubMed=9763455;
RA   Galcheva-Gargova Z., Gangwani L., Konstantinov K.N., Mikrut M.,
RA   Theroux S.J., Enoch T., Davis R.J.;
RT   "The cytoplasmic zinc finger protein ZPR1 accumulates in the nucleolus
RT   of proliferating cells.";
RL   Mol. Biol. Cell 9:2963-2971(1998).
CC   -!- FUNCTION: MAY BE A SIGNALING MOLECULE THAT COMMUNICATES MITOGENIC
CC       SIGNALS FROM THE CYTOPLASM TO THE NUCLEUS.
CC   -!- SUBUNIT: BINDS TO THE EGF AND PDGF RECEPTORS. BINDS TO THE
CC       ELONGATON FACTOR 1-ALPHA.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC; TRANSLOCATE TO THE NUCLEUS
CC       AFTER TREATMENT WITH MITOGENS.
CC   -!- SIMILARITY: BELONGS TO THE ZPR1 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
```

```
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; U41287; AAC52662.1; -.
DR   MGD; MGI:1330262; Zfp259.
DR   InterPro; IPR004457; Znf_ZPR1.
DR   Pfam; PF03367; ZPR1; 2.
DR   TIGRFAMs; TIGR00310; ZPR1_znf; 2.
KW   Nuclear protein; Zinc-finger.
FT   ZN_FING      51     83       C4-TYPE.
FT   ZN_FING     259    291       C4-TYPE.
SQ   SEQUENCE   459 AA;  50715 MW;  771D38C7B806044F CRC64;

  Query Match             6.9%;  Score 92.5;  DB 1;  Length 459;
  Best Local Similarity   22.9%;  Pred. No. 11;
  Matches   53;  Conservative   38;  Mismatches   81;  Indels   59;  Gaps   12;

Qy      33 PSGKKFRSKPQLARYLGNTVDLSSFDFRT---GKMMPSKLQKNKQR-----LRNDPLNQN 84
           |||  |   |  |    | : :: :| ||    :|:  : :  :::     |||: |  |
Db     199 PSGNSFVENPH-APQKDNALVITYYD-RTPQQAEMLGLQAEAPEEKAEEEDLRNEVLQFN 256

Qy      85 KGKPDLN----TTLPIRQTASIFKQPVTKVTN-----HPSNKVKS----DPQRMNEQPRQ 131
             |: |    | : : |    ||: :   ||     | :|:|||    :|
Db     257 TNCPECNAPAQTNMKLVQIPH-FKEVIIMATNCENCGHRTNEVKSGGAVEP--------- 306

Qy     132 LFWEKRLQGLSASDVTEQIIK----TMELPKGLQGVGPGSNDETLLSAVASALHTSSAPI 187
           |    |     ||:|  ::|    ::|:|:          :  | ||     |:   :
Db     307 LGTRITLHITDPSDMTRDLLKSETCSVEIPE---------LEFELGMAVLGGKFTTLEGL 357

Qy     188 TGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEAL 238
             :   | |||             | : |     | |::|:  :|| : :
Db     358 LKDIRELVTKNP-------------FTLGDSSNPDQSEKLQEFSQKLGQII 395


RESULT 11
NED1_MOUSE
ID   NED1_MOUSE     STANDARD;      PRT;   675 AA.
AC   P33215;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   NEDD1 protein (Fragment).
GN   NEDD1 OR NEDD-1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=92328780; PubMed=1378265;
RA   Kumar S., Tomooka Y., Noda M.;
RT   "Identification of a set of genes with developmentally down-regulated
RT   expression in the mouse brain.";
RL   Biochem. Biophys. Res. Commun. 185:1155-1161(1992).
CC   -!- FUNCTION: MAY PLAY AN IMPORTANT ROLE DURING THE EMBRYONIC
CC       DEVELOPMENT AND DIFFERENTIATION OF THE CENTRAL NERVOUS SYSTEM.
CC   -!- DEVELOPMENTAL STAGE: DOWN-REGULATED DURING THE DEVELOPMENT OF
CC       BRAIN.
CC   -!- SIMILARITY: CONTAINS AT LEAST 8 WD REPEATS (TRP-ASP DOMAINS).
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; D10712; BAA01554.1; ALT_INIT.
DR   MGD; MGI:97293; Nedd1.
DR   InterPro; IPR001680; WD40.
DR   Pfam; PF00400; WD40; 7.
DR   SMART; SM00320; WD40; 6.
DR   PROSITE; PS00678; WD_REPEATS_1; 2.
DR   PROSITE; PS50082; WD_REPEATS_2; 1.
DR   PROSITE; PS50294; WD_REPEATS_REGION; 1.
KW   Repeat; WD repeat.
FT   NON_TER       1      1
FT   REPEAT        2     42       WD 1.
FT   REPEAT       47     86       WD 2.
FT   REPEAT       90    129       WD 3.
FT   REPEAT      132    171       WD 4.
FT   REPEAT      175    215       WD 5.
FT   REPEAT      219    259       WD 6.
FT   REPEAT      261    300       WD 7.
FT   REPEAT      304    347       WD 8.
SQ   SEQUENCE   675 AA;  72969 MW;  B6D7D8DF21C31621 CRC64;

  Query Match             6.9%;  Score 92.5;  DB 1;  Length 675;
  Best Local Similarity   19.7%;  Pred. No. 18;
  Matches   55;  Conservative   53;  Mismatches  100;  Indels   71;  Gaps   13;

Qy      18 KSGLSAGKS-DVYYFSPSGKKFRSKPQLARYLGNTVDLSSFDFRTGKMMPSKLQKNKQRL 76
           ::||:  || |::           ||   |   | : | |||  |||   : |      :
Db     386 EAGLARSKSTDIF----------SKETDA---GKSQDFSSFD-DTGK---NSLGDMFSPI 428

Qy      77 RNDPL-----NQNKGK-------PDLNTTLPIRQTAS-----IFKQPVTKVTNHPSNKVK 119
           |:| :     ::: ||       | ||:  |:|: |      :     |  | | :
Db     429 RDDAVVSKGGDESIGKGDGLDFLPQLNSVFPLRKNAGASSSLVLHSSPLNVLMGSSGKEE 488

Qy     120 SDPQRMNEQPRQLFWEKRLQGLSASDVTEQIIKTME--LPKGLQGVGPGSND-------- 169
           ::   :: : :: :  |:       |  :|  | :      |:    | ||
Db     489 NESHELSAESKRAYLGKQ----EPKDAMKQFTKLISSGAEPGILNTCPSSNQARNLEKFE 544

Qy     170 ------------ETLLSAVASALHTSSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTD 217
                       |  ::  :: :  ::: :|  |:::: :     |  |:|      :|
Db     545 KPEKDIEAQLIHEPSVNGSSTTVPKAASSVTAGVASSLSEKIVDTLGNSRP---GAPLTS 601

Qy     218 EDIRKQEERVQQVRKKLEEALMADILSRAADTEEMDIEM 256
             ||  :  :|:      ||   ||::       : :||
Db     602 VQIRFIQNMIQETLDDFREACHRDIVN-------LQVEM 633


RESULT 12
MEFA_MOUSE
ID   MEFA_MOUSE     STANDARD;      PRT;   498 AA.
AC   Q60929;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   Myocyte-specific enhancer factor 2A.
GN   MEF2A.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Cerebellum;
RX   MEDLINE=97165895; PubMed=9013788;
RA   Lin X., Shah S., Bulleit R.F.;
RT   "The expression of MEF2 genes is implicated in CNS neuronal
RT   differentiation.";
RL   Brain Res. Mol. Brain Res. 42:307-316(1996).
CC   -!- FUNCTION: TRANSCRIPTION FACTOR WHICH BINDS SPECIFICALLY TO THE
CC       MEF2 ELEMENT PRESENT IN THE REGULATORY REGIONS OF MANY MUSCLE-
CC       SPECIFIC GENES. ACTIVATES TRANSCRIPTION VIA THIS ELEMENT. MAY BE
CC       INVOLVED IN MUSCLE-SPECIFIC AND/OR GROWTH FACTOR-RELATED
CC       TRANSCRIPTION (BY SIMILARITY).
CC   -!- SUBUNIT: HETERODIMER.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: BELONGS TO THE MADS DOMAIN FAMILY OF TRANSCRIPTION
CC       FACTORS. MEF2 SUBFAMILY.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; U30823; AAA74030.1; -.
DR   HSSP; P11831; 1SRS.
DR   TRANSFAC; T00505; -.
DR   MGD; MGI:99532; Mef2a.
DR   InterPro; IPR002100; TF_MADSbox.
DR   Pfam; PF00319; SRF-TF; 1.
DR   PRINTS; PR00404; MADSDOMAIN.
DR   SMART; SM00432; MADS; 1.
DR   PROSITE; PS00350; MADS_BOX_1; 1.
DR   PROSITE; PS50066; MADS_BOX_2; 1.
KW   Transcription regulation; Nuclear protein; DNA-binding; Activator;
KW   Multigene family.
FT   DOMAIN        3     57       MADS.
FT   DNA_BIND     58     86       MEF2-TYPE (POTENTIAL).
FT   DOMAIN      254    257       POLY-PRO.
FT   DOMAIN      288    293       POLY-GLU.
FT   DOMAIN      419    423       POLY-HIS.
FT   DOMAIN      448    455       POLY-SER.
SQ   SEQUENCE   498 AA;  53724 MW;  590678D1BD1B3723 CRC64;

  Query Match             6.8%;  Score 92;  DB 1;  Length 498;
  Best Local Similarity   27.4%;  Pred. No. 13;
  Matches   51;  Conservative   18;  Mismatches   51;  Indels   66;  Gaps   11;

Qy      32 SPSGKKF---RSKPQLARYLGNTVDLSSFDFRTGKMMPSKLQKNKQRLRNDP------LN 82
           || |  |   |: | |   :|||    |      ||:||:|         : |      |
Db     221 SPVGNGFVNSRASPNL---IGNTGANS-----LGKVMPTK---------SPPPPGGGSLG 263

Qy      83 QNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKVKSDPQRMNEQPRQLFWEKRLQGLS 142
             |  ||||   :|                  ||:|    |    |:      |   | :|
Db     264 MNSRKPDLRVAIP-----------------PSSKGMMPPLSEEEE-----LELNAQRIS 300

Qy     143 ASDVTEQI------IKTMELPKGLQGV------GPGSNDETLLSAVASALHTSSAP---I 187
           :|  |: :      : |  ||   ||:         : | :| ||  |||   ::|
Db     301 SSQATQPLATPVVSVTTPSLPP--QGLVYSAMPTAYNTDYSLTSADLSALQGFTSPGMLS 358

Qy     188 TGQVSA 193
           || ||
Db     359 LGQASA 364


RESULT 13
MAPA_RAT
ID   MAPA_RAT       STANDARD;      PRT;  2774 AA.
AC   P34926;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Microtubule-associated protein 1A (MAP 1A) [Contains: MAP1 light chain
DE   LC2].
GN   MAP1A.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=92355629; PubMed=1379599;
RA   Langkopf A., Hammarback J.A., Mueller R., Vallee R.B., Garner C.C.;
RT   "Microtubule-associated proteins 1A and LC2. Two proteins encoded in
RT   one messenger RNA.";
RL   J. Biol. Chem. 267:16561-16566(1992).
CC   -!- FUNCTION: Structural protein involved in the filamentous cross-
CC       bridging between microtubules and other skeletal elements.
CC   -!- SUBUNIT: 3 different light chains, LC1, LC2 and LC3, can associate
CC       with MAP1A and MAP1B proteins.
CC   -!- TISSUE SPECIFICITY: BRAIN, HEART AND MUSCLE.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED LATE DURING NEURONAL DEVELOPMENT
CC       APPEARING WHEN AXONS AND DENDRITES BEGIN TO SOLIDIFY AND STABILIZE
CC       THEIR MORPHOLOGY.
CC   -!- DOMAIN: THE BASIC REGION CONTAINING THE REPEATS MAY BE RESPONSIBLE
CC       FOR THE BINDING OF MAP1A TO MICROTUBULES.
CC   -!- PTM: VARIOUS SERINE RESIDUES MAY BE PHOSPHORYLATED BY CAMP KINASE.
CC   -!- PTM: LC2 IS COEXPRESSED WITH MAP1A. IT IS A POLYPEPTIDE GENERATED
CC       FROM MAP1A BY PROTEOLYTIC PROCESSING. IT IS FREE TO ASSOCIATE WITH
CC       BOTH MAP1A AND MAP1B.
CC   -!- SIMILARITY: TO MAP1B.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; M83196; AAB48069.1; -.
DR   PIR; A43359; A43359.
KW   Microtubules; Repeat; Phosphorylation.
FT   CHAIN     ?2465   2774       MAP1 LIGHT CHAIN LC2.
FT   DOMAIN      309    496       LYS-RICH (BASIC).
FT   DOMAIN      336    541       11 X 3 AA REPEATS OF K-K-[DE].
FT   REPEAT      336    338       1.
FT   REPEAT      415    417       2.
FT   REPEAT      420    422       3.
FT   REPEAT      424    426       4.
FT   REPEAT      427    429       5.
FT   REPEAT      431    433       6.
FT   REPEAT      436    438       7.
FT   REPEAT      440    442       8.
FT   REPEAT      444    446       9.
FT   REPEAT      449    451       10.
FT   REPEAT      539    541       11.
SQ   SEQUENCE   2774 AA;  299526 MW;  3DEF74427BA9D7D7 CRC64;

  Query Match             6.8%;  Score 92;  DB 1;  Length 2774;
  Best Local Similarity   22.4%;  Pred. No. 1.2e+02;
  Matches   63;  Conservative   46;  Mismatches  100;  Indels   72;  Gaps   15;

Qy      11 KKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNTVDLSSFDFRTGKMM----- 65
           ||||  ::    | | :        |:  :||:: :   :  ||  |     | :
Db     431 KKEEGRKEEKKDAKKDE--------KRKDTKPEVKKL--SKPDLKPFTPEVRKTLYKAKA 480

Qy      66 PSKLQKNKQRLRNDPLNQNKGKPDLNT---TLPIRQTASIFKQPVTKVTNHPSNKVKS-- 120
           | ::: :| |          :|: :|::   | | :: |:    |   |: |    : |
Db     481 PGRVKVDKGRAA-------RGEKELSSEPRTPPAQKGAA----PPAAVSGHRELALSSPE 529

Qy     121 ----DPQRMNEQPRQLFWEKRLQGLS----ASDVTEQ-----IIKTMELPKG--LQGVGP 165
               | : :  : | |  |:|  ||      :| |||      |:  : | |  |:|
Db     530 DLTQDFEELKREERGLLAEQRDTGLGEKPLPADATEQGHPSAAIQVTQ-PSGPVLEGEHV 588

Qy     166 GSNDETLLSAVASALHTSSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEE 225
           | :  :      |:  | :|   | |||    |                  : ||| |
Db     589 EREKEVVPDSPGDKGSTNRGPDSG---AEVEKEKETW---------------EERKQRE 629

Qy     226 -----RVQQVRKKLEEALMADILSRAADTEEMDIEMDSGDE 261
                |:: |  :  |:: : |: |||:    | :|
Db     630 AELGPENTAAREESEAEVKEDVIEK-AELEEMEETHPSDEE 669


RESULT 14
AROA_PSES2
ID   AROA_PSES2     STANDARD;      PRT;   449 AA.
AC   P56952;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   3-phosphoshikimate 1-carboxyvinyltransferase (EC 2.5.1.19) (5-
```

```
DE   enolpyruvylshikimate-3-phosphate synthase) (EPSP synthase) (EPSPS).
GN   AROA.
OS   Pseudomonas sp. (strain PG2982), and
OS   Achromobacter sp. (strain LBAA).
OC   Bacteria; Proteobacteria.
OX   NCBI_TaxID=308, 129026;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 2-16.
RA   Barry G.F., Kishore G.M., Padgette S.R., Stallings W.C.;
RT   "Glyphosate-tolerant 5-enolpyruvylshikimate-3-phosphate synthases.";
RL   Patent number US5633435, 27-MAY-1997.
CC   -!- CATALYTIC ACTIVITY: Phosphoenolpyruvate + 3-phosphoshikimate =
CC       phosphate + 5-O-(1-carboxyvinyl)-3-phosphoshikimate.
CC   -!- PATHWAY: Aromatic amino acids biosynthesis; shikimate pathway;
CC       sixth step.
CC   -!- SUBUNIT: MONOMER (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC   -!- MISCELLANEOUS: RESISTANT TO THE ANTIBIOTIC GLYPHOSATE.
CC   -!- SIMILARITY: BELONGS TO THE EPSP SYNTHASE FAMILY.
DR   InterPro; IPR001986; EPSP_syntase.
DR   Pfam; PF00275; EPSP_syntase; 1.
DR   ProDom; PD001867; EPSP_syntase; 1.
DR   PROSITE; PS00104; EPSP_SYNTHASE_1; 1.
DR   PROSITE; PS00885; EPSP_SYNTHASE_2; 1.
KW   Aromatic amino acid biosynthesis; Transferase; Herbicide resistance.
SQ   SEQUENCE   449 AA;  47297 MW;  447F213EECCAEFC1 CRC64;

  Query Match             6.7%;  Score 90.5;  DB 1;  Length 449;
  Best Local Similarity   22.0%;  Pred. No. 15;
  Matches   57;  Conservative   38;  Mismatches  123;  Indels   41;  Gaps    9;

Qy      16 IRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGN-------TVDL-SSFDFRTGKMMPS 67
           ||| |      ||: : |     :|: |   ||       |: |  ::| :|  :  :
Db      70 IRKEG------DVWIINGVGNGCLLQPEAALDFGNAGTGARLTMGLVGTYDMKTSFIGDA 123

Qy      68 KLQKNKQRLRNDPLNQNKGKPDL--NTTLPIRQTASIFKQPVTKVTNHPSNKVKSDPQRM 125
             | |       :|| :   : :      :|:       |:|      | :|||
Db     124 SLSKRPMGRVLNPLREMGVQVEAADGDRMPLTLIGPKTANPITYRVPMASAQVKS----- 178

Qy     126 NEQPRQLFWEKRLQGLSASDVT---EQIIKTMELPKGLQGVGPGSNDETLLSAVASALHT 182
                  | ||:   ||   | ::      | ||| |     ||   |     |
Db     179 ---------AVLLAGLNTPGVTTVIEPVMTRDHTEKMLQGFGADLTVETDKDGVRHIRIT 229

Qy     183 SSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEALMADI 242
             : || :  |  :|:   :|: ||  | :|   |:  :   :   |      |:  :
Db     230 GQGKLVGQ-TIDVPGDPS---STAFPLVAALLVEGSDVTIRNVLMNPTR----TGLILTL 281

Qy     243 LSRAADTEEMDIEMDSGDE 261
             || | ::  :  |::
Db     282 QEMGADIEVLNARLAGGED 300


RESULT 15
Y364_RICPR
ID   Y364_RICPR     STANDARD;      PRT;   386 AA.
AC   Q9ZDG5;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein RP364.
GN   RP364.
OS   Rickettsia prowazekii.
OC   Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC   Rickettsiaceae; Rickettsieae; Rickettsia.
OX   NCBI_TaxID=782;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Madrid E;
RX   MEDLINE=99039499; PubMed=9823893;
RA   Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA   Sicheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA   Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT   "The genome sequence of Rickettsia prowazekii and the origin of
RT   mitochondria.";
RL   Nature 396:133-140(1998).
CC   -!- SIMILARITY: SOME, TO R.PROWAZEKII RP363.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AJ235271; CAA14823.1; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   386 AA;  44346 MW;  AAEE08EB8D43A0E7 CRC64;

  Query Match             6.7%;  Score 90;  DB 1;  Length 386;
  Best Local Similarity   21.3%;  Pred. No. 13;
  Matches   50;  Conservative   45;  Mismatches   74;  Indels   66;  Gaps   11;

Qy       8 PGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNTVDLSSFDFRTGKMMPS 67
           | ::: :: :| || : : :     |:||          |:   :|      : |  :|
Db     170 PLFRESDIYKKLGLKSAEIEREIQDPNGK----------YVQQLIDA-----KIGSNIPF 214

Qy      68 KLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKVKSDPQRMNE 127
             :|||         | |:||        | :|| | :: :||       |   : :: :|
Db     215 HMQKN---------NVNEGK-------EIERTA-IIEKAITKF--EQDKKFSFEGKKRDE 255

Qy     128 QPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSAVASALHTSSAPI 187
             : |   | |:|  |||      |   :    ||:  |   :|| | ||: :   |  |
Db     256 ITKYL--SKSLEG--ASDYILTFKKNELVDVIYQGLDKG---QTLWSKVANYIGIKSYSI 308

Qy     188 TGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQ--------EERVQQVRKKL 234
           : :   :| |                : :: |:          :::::|: |:|
Db     309 SKENLKSVAK----------------IINDKIKSSHTPLKIEVQDKLKQISKEL 346


Search completed: March 12, 2003, 08:44:15
Job time : 17.5111 secs
```

OM protein - protein search, using sw model

Run on: March 12, 2003, 03:22:03 ; Search time 19.0758 Seconds
(without alignments)
1320.377 Million cell updates/sec

Title: US-09-554-414B-2_COPY_150_411
Perfect score: 1344
Sequence: 1 MDCPALPPGWKKEEVIRKSG.....LSRAADTEEMDIEMDSGDEA 262

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	177	13.2	476	2	S57963	methyl CpG binding
2	167	12.4	492	2	A41907	methyl-CpG-binding
3	118	8.8	186	2	T48092	hypothetical prote
4	117.5	8.7	384	2	G86287	hypothetical prote
5	115	8.6	820	2	T12972	hypothetical prote
6	109.5	8.1	182	2	T45595	hypothetical prote
7	107.5	8.0	955	2	S52959	male-specific leth
8	105.5	7.8	527	2	B64633	hypothetical prote
9	104.5	7.8	912	2	T01769	hypothetical prote
10	104.5	7.8	1430	2	T34516	hypothetical prote
11	103	7.7	566	2	T04569	hypothetical prote
12	101	7.5	792	2	E84525	hypothetical prote
13	100	7.4	375	2	E81442	probable MCP-domai
14	99	7.4	789	2	E84514	hypothetical prote
15	99	7.4	1009	2	S49618	helicase-like tran
16	98.5	7.3	649	2	T37740	coiled-coil protei
17	97	7.2	448	2	I50730	yes-associated pro
18	97	7.2	684	2	C84434	hypothetical prote
19	95	7.1	794	2	T37989	DNA mismatch repai
20	93.5	7.0	1024	2	T30868	RhoA-binding prote
21	93	6.9	1048	2	S64758	SCD25 protein (ver
22	93	6.9	5762	2	A41819	proline-rich pepti
23	92.5	6.9	660	2	I60167	regulatory protein
24	92	6.8	345	2	S77631	probable integrase
25	92	6.8	493	2	G89876	bifunctional purin
26	92	6.8	1560	2	T42727	proliferation pote
27	92	6.8	1746	2	D83181	hypothetical prote
28	92	6.8	2774	2	A43359	microtubule-associ
29	91.5	6.8	611	2	T14738	hypothetical prote
30	91	6.8	345	2	S77632	probable integrase
31	91	6.8	357	2	T48747	hypothetical prote
32	91	6.8	643	2	G87626	DNA primase [impor
33	91	6.8	1250	2	S14177	SCD25 protein (ver
34	91	6.8	1610	2	D89451	protein T04G9.1 [i
35	90.5	6.7	469	2	T46929	hypothetical prote
36	90.5	6.7	469	2	T46930	hypothetical prote
37	90.5	6.7	496	2	T15691	hypothetical prote
38	90.5	6.7	622	2	C84506	probable Athila re
39	90.5	6.7	658	2	E71465	hypothetical prote
40	90.5	6.7	934	2	T47546	protein kinase-lik
41	90.5	6.7	990	2	I51618	nucleolar phosphop
42	90.5	6.7	1166	2	T13958	synGAP-b1 protein
43	90.5	6.7	1249	2	T14270	Ras-GTPase activat
44	90.5	6.7	1293	2	T14259	ras GTPase-activat
45	90	6.7	345	2	A89991	integrase [importe

ALIGNMENTS

RESULT 1
S57963
methyl CpG binding protein 2 - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 05-Nov-1999
C;Accession: S57963
R;d'Esposito, M.; Quaderi, N.A.; Ciccodicola, A.; Bruni, P.; Esposito, T.; D'Urso, M.
submitted to the EMBL Data Library, July 1995
A;Description: Physical mapping and expression analysis of an X-linked gene encoding
A;Reference number: S57963
A;Accession: S57963
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-476 <DES>
A;Cross-references: EMBL:X89430; NID:g899295; PIDN:CAA61599.1; PID:g899296

Query Match 13.2%; Score 177; DB 2; Length 476;
Best Local Similarity 28.4%; Pred. No. 1e-05;
Matches 75; Conservative 37; Mismatches 86; Indels 66; Gaps 14;

```
Qy       2 DCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARY---LGNT-VDLSSF 57
           | | || || ::   |||| |||| |||  :| || |||| :|  |   :|:| :| : |
Db      86 DDPTLPEGWTRKLKQRKSGRSAGKYDVYLINPQGKAFRSKVELIAYFEKVGDTSLDPNDF 145

Qy      58 DFR-TGKMMPSKLQKNKQRLRNDP----LNQNKGKP-DLNTTLPIRQTASIFKQPVTKVT 111
           ||  ||:  ||: ::   :    |       : :|:|    || |   |:   :  | :|
Db     146 DFTVTGRGSPSRREQKPPKKPKSPKAPGTGRGRGRPKGSGTTRPKAATSEGVQ--VKRVL 203

Qy     112 NHPSNKVKSDPQRMNEQPRQLFWEKRLQGLSASDVTE-QIIK-----------TMELPKG 159
                ||  : : : | |     : :|  |:  |:  :||              :||
Db     204 E------KSPGKLLVKMPFQTSPGGKAEGGGATTSTQVMVIKRPGRKRKAEADPQAIPK- 256

Qy     160 LQGVGPGSNDETLLSAVASALHTSSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDED 219
           :|  ||        | ||:|             ||  |  |              | :
Db     257 KRGRKPG-------SVVAAA-------------AAEAKKKA--------------VKESS 282

Qy     220 IRKQEERVQQVRK-KLEEALMADI 242
           ||  :| |  ::| |  | :  ::
Db     283 IRSVQETVLPIKKRKTRETVSIEV 306
```

RESULT 2
A41907
methyl-CpG-binding protein 2 - rat
N;Alternate names: chromosomal protein MeCP2
C;Species: Rattus norvegicus (Norway rat)
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 08-Oct-1999
C;Accession: A41907; S41461
R;Lewis, J.D.; Meehan, R.R.; Henzel, W.J.; Maurer-Fogy, I.; Jeppesen, P.; Klein, F.;
Cell 69, 905-914, 1992
A;Title: Purification, sequence, and cellular localization of a novel chromosomal pro
A;Reference number: A41907; MUID:92298389; PMID:1606614

A;Accession: A41907
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-492 <LEW>
A;Cross-references: GB:M94064; NID:g205361; PIDN:AAA41584.1; PID:g205362
R;Nan, X.; Meehan, R.R.; Bird, A.
Nucleic Acids Res. 21, 4886-4892, 1993
A;Title: Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2.
A;Reference number: S41461; MUID:94232813; PMID:8177735
A;Contents: annotation; methyl CpG-binding domain
C;Keywords: chromosomal protein; DNA binding
F;78-162/Domain: methyl-CpG-binding #status experimental <MCG>

Query Match 12.4%; Score 167; DB 2; Length 492;
Best Local Similarity 28.1%; Pred. No. 5.9e-05;
Matches 74; Conservative 37; Mismatches 88; Indels 64; Gaps 13;

```
Qy       2 DCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARY---LGNT-VDLSSF 57
Db      96 DDPTLPEGWTRKLKQRKSGRSAGKYDVYLINPQGKAFRSKVELIAYFEKVGDTSLDPNDF 155

Qy      58 DFR-TGKMMPSKLQKNKQRLRNDP----LNQNKGKPDLNTTLPIRQTASIFKQPVTKVTN 112
Db     156 DFTVTGRGSPSRREQKPPKKPKSPKAPGTGRGRGRPKGSGTGRPKAAASEGVQ-VKRVLE 214

Qy     113 HPSNKVKSDPQRMNEQPRQLFWEKRLQGLSASDVTE-QIIK----------TMELPKGL 160
Db     215 ------KSPGKLLVKMPFQASPGGKGEGGGATTSAQVMVIKRPGRKRKAEADPQAIPK-K 267

Qy     161 QGVGPGSNDETLLSAVASALHTSSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDI 220
Db     268 RGRKPG-------SVVAAA-------------AAEAKKKA--------------VKESSI 293

Qy     221 RKQEERVQQVRK-KLEEALMADI 242
Db     294 RSVQETVLPIKKRKTRETVSIEV 316
```

RESULT 3
T48092
hypothetical protein T20O10.130 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C;Accession: T48092
R;Obermaier, B.; Ottenwaelder, B.; Duchemin, D.; Zeitler, K.; Mewes, H.W.; Rudd, S.; Lem
submitted to the Protein Sequence Database, April 2000
A;Reference number: Z24484
A;Accession: T48092
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-186 <OBE>
A;Cross-references: EMBL:AL163816
A;Experimental source: cultivar Columbia; BAC clone T20O10
C;Genetics:
A;Map position: 3
A;Introns: 16/3
A;Note: T20O10.130

Query Match 8.8%; Score 118; DB 2; Length 186;
Best Local Similarity 28.8%; Pred. No. 0.08;
Matches 30; Conservative 18; Mismatches 38; Indels 18; Gaps 3;

```
Qy       4 PALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLG-----NTVDLSSFD 58
Db      91 PKTPKGFKRSLVLRK---DYSKMDTYYFTPTGKKLRSRNEIAAFVEANPEFRNAPLGDFN 147

Qy      59 FRTGKMMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTASI 102
Db     148 FTVPKVMEDTV----------PPDPKLGSPFPSTTTTTSEKSSV 181
```

RESULT 4
G86287

hypothetical protein F9L1.28 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Dec-2001
C;Accession: G86287
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: G86287
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-384 <STO>
A;Cross-references: GB:AE005172; NID:g5103831; PIDN:AAD39661.1; GSPDB:GN00141
C;Genetics:
A;Map position: 1

Query Match 8.7%; Score 117.5; DB 2; Length 384;
Best Local Similarity 20.2%; Pred. No. 0.23;
Matches 60; Conservative 61; Mismatches 137; Indels 39; Gaps 13;

```
Qy       1 MDCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYL----GNTVDLSS 56
Db      10 IELPA-PASWKKLFYPKRAG-TPRKTEIVFVAPTGEEISSRKQLEQYLKAHPGNPV-ISE 66

Qy      57 FDFRTGKMMPSKLQKNKQRLR-------NDPL----------NQNKGKPDLNTTLPIRQT 99
Db      67 FEWTTGE-TPRRSSRISQKVKATTPTPDKEPLLKKRRSSLTKKDNKEAAEKNEEAAVKEN 125

Qy     100 ASIFKQPVTKVTNHPSNKVK------SDPQRMNEQPRQLFWEK-RLQGLSASDVTEQIIK 152
Db     126 MDVDKDGKTENAEAEKEKEKEGVTEIAEAEKENNEGEKTEAEKVNKEGEKTEAGKEGQTE 185

Qy     153 TMELPKGLQG-VGPGSNDET-LLSAVASALHTSSAPI---TGQVSAAVEKNPAVWLNTSQ 207
Db     186 IAEAEKEKEGEKAEAENKEAEVVRDKKESMEVDTSELEKKAGSGEGAEEPSKVEGLKDTE 245

Qy     208 PLCKAFIVTDEDIRKQ--EERVQQVRKKLEEALMADILSRAADTEEMDIEMDSGDEA 262
Db     246 MKEAQEVVTEADVEKKPAEEKTENKGSVTTEANGEQNVTLGEPNLDADAEADKGKES 302
```

RESULT 5
T12972
hypothetical protein T21L8.10 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 18-Aug-2000
C;Accession: T12972
R;Choisne, N.; Robert, C.; Brottier, P.; Wincker, P.; Cattolico, L.; Artiguenave, F.;
submitted to the Protein Sequence Database, July 1999
A;Reference number: Z17586
A;Accession: T12972
A;Molecule type: DNA
A;Residues: 1-820 <CHO>
A;Cross-references: EMBL:AL096860; GSPDB:GN00061; ATSP:T21L8.10
A;Experimental source: cultivar Columbia; BAC clone T21L8
C;Genetics:
A;Gene: ATSP:T21L8.10
A;Map position: 3
A;Introns: 115/1; 297/3; 431/3; 443/3; 484/3; 520/3; 539/3; 560/3; 641/2; 668/3; 726/
C;Superfamily: Arabidopsis thaliana hypothetical protein T21L8.10

Query Match 8.6%; Score 115; DB 2; Length 820;
Best Local Similarity 23.1%; Pred. No. 0.97;
Matches 65; Conservative 37; Mismatches 99; Indels 80; Gaps 14;

```
Qy      18 KSGLSAGKSDVYYFSPS---GKKFRSKPQLARYLGNTVDLSSFDFRTGKMMPSKLQK--- 71
```

```
Db     511 KKRVSRGRNENKRVKPSVHAEDNFKTRKQMPRKRQKQVDSADVDIPTRKEAQSKKRKIIG 570

Qy      72 -NKQRLRND-------PLNQNKGK----PDLNTTLPIRQTASIFKQPVTKVTNHPSNKVK 119
Db     571 NDGDNADNDGDNDDFQPAPQRKSKRGTVPSIHTQAPF--TAEKKKHPIL----HPFAKV- 623

Qy     120 SDPQRMNEQPRQLFWEKRLQ----GLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSA 175
Db     624 -DATRLEKLAE---WKKSRKNKPLSIAGNNVDTKWFTTLETP----------------GK 663

Qy     176 VASALHTSSAPITGQVSAAVE-------------KNPAV------WLNTSQPLCKAFIVT 216
Db     664 AITATH---------VDAALELMKTRKESNPELFKNKSVVFVGSSFLNVIDESYMEFLDN 714

Qy     217 DEDIRKQEERVQQVRKKLEEA---LMADILSRAADTEEMDI 254
Db     715 KEGFQFQSEEIKEVKKKYVAAYSIAMPYIVRNILKKEDMDV 755


RESULT 6
T45595
hypothetical protein F12A12.100 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C;Accession: T45595
R;Choisne, N.; Robert, C.; Brottier, P.; Wincker, P.; Cattolico, L.; Artiguenave, F.; Sa
submitted to the Protein Sequence Database, December 1999
A;Reference number: Z23008
A;Accession: T45595
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-182 <CHO>
A;Cross-references: EMBL:AL133314
A;Experimental source: cultivar Columbia; BAC clone F12A12
C;Genetics:
A;Map position: 3
A;Introns: 57/3; 99/3
A;Note: F12A12.100

  Query Match             8.1%;  Score 109.5;  DB 2;  Length 182;
  Best Local Similarity   27.7%;  Pred. No. 0.34;
  Matches   39;  Conservative   20;  Mismatches   51;  Indels   31;  Gaps    7;

Qy       6 LPPGWKKEEVIRKSGLSAGKSDVYYFSP-SGKKFRSKPQLARYLGNTVDLSSFDFRTGKM 64
Db      35 LPPDWRTEIRVRTSGTKAGTVDKFYYEPITGRKFRSKNEVLYYL---------EHGTPKK 85

Qy      65 MPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKVKSDPQR 124
Db      86 KSVKTAEN----GDSHSEHSEGRGSAR-----RQTKS-----NKKVTEPPPKPLNFD--- 128

Qy     125 MNEQPRQLFWEKRLQGLSASD 145
Db     129 FLNVPEKVTW----TGINGSE 145


RESULT 7
S52959
male-specific lethal-1 protein - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 19-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 16-Feb-1997
C;Accession: S52959; S65350
R;Palmer, M.J.; Mergner, V.A.; Richman, R.; Manning, J.E.; Kuroda, M.I.; Lucchesi, J.C.
Genetics 134, 545-557, 1993
A;Title: The male-specific lethal-one (msl-1) gene of Drosophila melanogaster encodes a
A;Reference number: S52959; MUID:93314941; PMID:8325488
A;Accession: S52959
A;Molecule type: DNA
A;Residues: 1-955 <PAL>
A;Cross-references: EMBL:L14582
R;Palmer, M.J.; Mergner, V.A.; Richman, R.; Manning, J.E.; Kuroda, M.I.; Lucchesi, J.C.
submitted to the EMBL Data Library, April 1993
A;Reference number: S65350
A;Accession: S65350
A;Molecule type: DNA
A;Residues: 1-407,'S',409-585,'M',587-955 <PAW>
A;Cross-references: EMBL:L14582
C;Genetics:
A;Gene: msl-1
A;Cross-references: FlyBase:FBgn0005617
A;Introns: 329/3

  Query Match             8.0%;  Score 107.5;  DB 2;  Length 955;
  Best Local Similarity   23.6%;  Pred. No. 4.4;
  Matches   48;  Conservative   26;  Mismatches   58;  Indels   71;  Gaps    7;

Qy      73 KQRLRNDPLNQNKGKP----------DLNTTLPIRQTASIFKQPVTKV-TNHPSNKVKS 120
Db     474 KETLKRQPEDAPKHLPKAVAPKVTKTSSRESTLPKANTADIKDAPAQKVIANHQSTKTQT 533

Qy     121 DP---QRMNEQPRQ--LFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDET---- 171
Db     534 DPVKTQRLQVKIRQYEMHPDMRTGSSAPSDIRKQ----------KNVDPVSTPETKTIK 582

Qy     172 ----LLSAVASALHTSSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERV 227
Db     583 SKSILVNDKKTTSETSQSP---------------------------------DQEIDV 607

Qy     228 QQVRKKLEEALMADILSRAADTE 250
Db     608 ETVRRKLAEHLKKELLSQSHSSQ 630


RESULT 8
B64633
hypothetical protein HP0906 - Helicobacter pylori (strain 26695)
C;Species: Helicobacter pylori
C;Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999
C;Accession: B64633
R;Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R
 Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKe
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey,
Nature 388, 539-547, 1997
A;Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser,
A;Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A;Reference number: A64520; MUID:97394467; PMID:9252185
A;Accession: B64633
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-527 <TOM>
A;Cross-references: GB:AE000600; GB:AE000511; NID:g2314042; PIDN:AAD07958.1; PID:g231

  Query Match             7.8%;  Score 105.5;  DB 2;  Length 527;
  Best Local Similarity   27.6%;  Pred. No. 2.8;
  Matches   48;  Conservative   26;  Mismatches   75;  Indels   25;  Gaps    8;

Qy      78 NDPLNQNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKVKSD--PQRMNEQPRQLFWE 135
Db     133 NETLNKNEKKPN-GVTSSVHQTNLTNKNPITP-TNHANNAIKNPTAPTDTKKEPKTL--- 187

Qy     136 KRLQGLSAS-DVTEQIIKTMELPKGLQGVGPGSNDETLLSAVASALHTSSAPITGQVSAA 194
Db     188 KDIQTLSQKHDLNASNIQAATTPE---------NKNPLNASDQLALKTTQTPTNHTLAKN 238

Qy     195 VEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKK---LEEALMADILSR 245
Db     239 DAKNTANLSSVLQSLEK-----KEPQNKEHANPLNNEKKTPPLKEALEMNAIKR 287


RESULT 9
T01769
hypothetical protein A_IG002P16.18 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 19-Feb-1999 #sequence_revision 19-Feb-1999 #text_change 15-Sep-2000
C;Accession: T01769
```

```
R;Miller, N.; Beck, C.; Kramer, J.
submitted to the EMBL Data Library, June 1997
A;Description: The sequence of A. thaliana IG002P16.
A;Reference number: Z14421
A;Accession: T01769
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-912 <MIL>
A;Cross-references: EMBL:AF007270; NID:g2191157; PID:g2191167; GSPDB:GN00063; ATSP:A_IG
C;Genetics:
A;Gene: ATSP:A_IG002P16.18
A;Map position: 5
A;Introns: 37/1; 71/2; 112/3; 648/3; 686/1
C;Superfamily: Arabidopsis thaliana hypothetical protein F14P22.240

  Query Match             7.8%;  Score 104.5;  DB 2;  Length 912;
  Best Local Similarity   18.8%;  Pred. No. 6.9;
  Matches   58;  Conservative   48;  Mismatches  101;  Indels  101;  Gaps   12;

Qy      11 KKEEVIRKSGLSAGKSDVYYFS-PSGKKFRSK-------PQLARYLGNTVDLSSFDFRTG 62
           :| |  :|        :| :   ||||    :       |  ::::|     || |   |
Db     302 EKLEAAQKVSSRQNSNDTFNLKYPSGKHNEKRITTSLTTPSTSKFMGK----SSTDGLKG 357

Qy      63 KMMPSKLQ-----------------------------KNKQRLRNDPLN-------QNK 85
           |: || :                              |::  ||  |::         ||
Db     358 KVKPSYVSAQAKAGTTPLSVTRNSANQKEKADAKKCVVKSQNALRGAPISMGKNMFKQNN 417

Qy      86 GKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKVKSDPQRMNEQPRQLFWEKRLQGLSASD 145
           | :     |     |:  |  :|| |   |||  :   :::|           |||  :
Db     418 QKQNCRDNQP--SMTSVLNQKSSKVNNKVVNKVPVESGSISKQ----------LGLSTAS 465

Qy     146 VTEQ----------IIKTMELPKGLQ-------------------------GVGPGSNDE 170
           :           : :: :|| |:|                         |:  | :|
Db     466 AEKNTSLSLSRKKTLPRSKKLPNGMQKSGISDDKRTKRSENMIKCNITIDGGLNKGKDDR 525

Qy     171 TLLSAVASALHTSSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRK-QEERVQQ 229
           | |   | |:|| |  | ::     :  :|   :  :| :  : :    |:::::
Db     526 KKEMDVIS--FTFSSPIKGLSSDSLSSTQGIGQDTDSAV--SFNIGGDSLNALLEQKLRE 581

Qy     230 VRKKLEEA 237
           :  ||| :
Db     582 LTSKLESS 589


RESULT 10
T34516
hypothetical protein ZK783.4 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 02-Sep-2000
C;Accession: T34516
R;Favello, A.; Vaudin, M.
submitted to the EMBL Data Library, August 1994
A;Description: The sequence of C. elegans cosmid ZK783.
A;Reference number: Z21536
A;Accession: T34516
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1430 <FAV>
A;Cross-references: EMBL:U13646; PIDN:AAC24421.1; GSPDB:GN00021; CESP:ZK783.4
A;Experimental source: strain Bristol N2; clone ZK783
C;Genetics:
A;Gene: CESP:ZK783.4
A;Map position: 3
A;Introns: 248/3; 373/3; 547/1; 593/2; 905/2; 1042/3; 1116/3; 1317/2; 1376/1
C;Superfamily: bromodomain homology
F;1284-1339/Domain: bromodomain homology <BRO>

  Query Match             7.8%;  Score 104.5;  DB 2;  Length 1430;
  Best Local Similarity   23.5%;  Pred. No. 13;
  Matches   38;  Conservative   33;  Mismatches   72;  Indels   19;  Gaps    6;

Qy       9 GWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNT----VDLSSFDFRTGKM 64
           +|:::  :|    :  | || ||:| |||  :  :: |||       :   :| | | |:

Db     336 GWRRQTCVRSIASAGVKGDVSYFAPCGKKLSTYSEVVRYLTKNSIHYITRDNFLFNT-KL 394

Qy      65 MPSKLQKNKQRLRNDPLNQNK----GKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKVKS 120
           :  :    ||   ::   : :     : |:|  |    |     : | |:    || |
Db     395 VIGEFIVPKQTEADETQQEREFAMFTEDDINKEL----TRLNVLKFVPKIQASTSNGVHE 450

Qy     121 DPQRMN--EQPRQLFWEKRLQGLSASDVTEQIIKTMELPKGL 160
           |  :|:  |:| :       |       : ||::: :  :  |:
Db     451 DDIKMSKIEEPDEPLDPSELN----DEFTEELVHSQIMSNGV 488


RESULT 11
T04569
hypothetical protein T12H17.130 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 14-May-1999
C;Accession: T04569
R;Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancro
submitted to the Protein Sequence Database, February 1998
A;Reference number: Z15377
A;Accession: T04569
A;Molecule type: DNA
A;Residues: 1-566 <BEV>
A;Cross-references: EMBL:AL021635
A;Experimental source: cultivar Columbia; BAC clone T12H17
C;Genetics:
A;Map position: 4
A;Introns: 34/3; 194/3; 349/1; 433/3; 459/1; 493/3; 505/1
A;Note: T12H17.130

  Query Match             7.7%;  Score 103;  DB 2;  Length 566;
  Best Local Similarity   32.8%;  Pred. No. 4.8;
  Matches   22;  Conservative   14;  Mismatches   23;  Indels    8;  Gaps    2;

Qy       4 PALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNTVD-----LSSFD 58
           |  | |:|:  ::||      | | || :|:||| :|: ::| ::     |     |  |:
Db     112 PRTPRGFKRSLILRK---DYSKMDAYYITPTGKKLKSRNEIAAFIDANQDYKYALLGDFN 168

Qy      59 FRTGKMM 65
           |   |:|
Db     169 FTVPKVM 175


RESULT 12
E84525
hypothetical protein At2g15140 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C;Accession: E84525
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: E84525
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-792 <STO>
A;Cross-references: GB:AE002093; NID:g4585915; PIDN:AAD25576.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g15140
A;Map position: 2
C;Superfamily: Arabidopsis thaliana hypothetical protein T21L8.10

  Query Match             7.5%;  Score 101;  DB 2;  Length 792;
  Best Local Similarity   22.8%;  Pred. No. 11;
  Matches   64;  Conservative   35;  Mismatches  102;  Indels   80;  Gaps   14;

Qy      18 KSGLSAGKSDVYYFSPS---GKKFRSKPQLARYLGNTVDLSSFDFRTGKMMPSKLQK--- 71
           |  :| |::       ||       ::: |: |     || :  |  | |   || :|
Db     483 KKRVSRGRNQNKRVKPSVHAEDNLKTRKQVPRKRQKQVDSADVDVPTRKEAESKKRKIIG 542
```

```
Qy      72 -NKQRLRND-------PLNQNKGK----PDLNTTLPIRQTASIFKQPVTKVTNHPSNKVK 119
           :     ||        |  | | |     | ::|  |   ||   | |:      ||  ||
Db     543 NDGDNADNDGDNDDFQPAPQRKSKRDTVPSIHTQAPF--TAEKKKHPIL----HPFAKV- 595

Qy     120 SDPQRMNEQPRQLFWEKRLQ----GLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSA 175
           |  |: :      |:|  :     :: : |  :    |:| |
Db     596 -DATRLEKLAE---WKKSRKNKPLSIAGNIVDTKWFTTLETP---------------GK 635

Qy     176 VASALHTSSAPITGQVSAAVE-------------KNPAV------WLNTSQPLCKAFIVT 216
             :| |          | ||:|              || :|      :||        |:
Db     636 AITATH---------VDAALELMKTRKESNPELFKNKSVVFVGSSFLNVIDESYMEFLDN 686

Qy     217 DEDIRKQEERVQQVRKKLEEA---LMADILSRAADTEEMDI 254
           |  : | | :::|:||   |     |  |:      |:||:
Db     687 KEGFQFQSEEIKEVKKKYVAAYSITMPYIVRNILKKEDMDV 727


RESULT 13
E81442
probable MCP-domain signal transduction protein Cj0246c [imported] - Campylobacter jejun
C;Species: Campylobacter jejuni
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002
C;Accession: E81442
R;Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chillin
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrel
Nature 403, 665-668, 2000
A;Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp
A;Reference number: A81250; MUID:20150912; PMID:10688204
A;Accession: E81442
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-375 <PAR>
A;Cross-references: GB:AL139074; GB:AL111168; NID:g6967505; PIDN:CAB72714.1; PID:g696772
A;Experimental source: serotype O2, strain NCTC 11168
C;Genetics:
A;Gene: Cj0246c

  Query Match             7.4%;  Score 100;  DB 2;  Length 375;
  Best Local Similarity   20.6%;  Pred. No. 4.6;
  Matches   54;  Conservative   47;  Mismatches  101;  Indels   60;  Gaps   10;

Qy      20 GLSAGKSDVYYFSPSGKKFRS-KPQLARYLGNTVD-LSSFDFRTGKMMPSKLQKNKQ--R 75
           |:   || |  |    : |   ||   |||    :   | | :    :   ||:   :  :
Db     105 GVGIAKSLVAIFFYEVENFNDFKP---RYLKTFIQKYSDFKYYYLNIRAQKLEMTNEINK 161

Qy      76 LRNDPLNQNKGKPDLNTTL---------PIRQTASIFKQPVTKVTNHPSNKVKSDPQRMN 126
           :  : | ||  : | ||::          ||:: :   :  |  ||:    :     ::||
Db     162 IILNQLKQNTSEIDKNTSIFKEIFEELENIRRSLTTISESFTNFTNYLEYNLYQSEEKMN 221

Qy     127 EQPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSAVASALHTSSAP 186
             || :|  |:    :|:   ::|  |:          ::|:| :: : :  : |
Db     222 -------LEKEVQ--SSLKNIDQLNSILDLISGIA-------EQTILLSLNAGIEAARAG 265

Qy     187 ITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEALMADI---- 242
           |                        : | |  :::||  |  |    ::| |:   |
Db     266 KLG--------------------RGFAVVADEVRKLSENTQMGLGEMEGAIKLVIQTIQ 304

Qy     243 ---LSRAADTEEMDIEMDSGDE 261
             |  : |:||:   |  :|
Db     305 SIAKSSNSSTQEMNFIRDKTNE 326


RESULT 14
E84514
hypothetical protein At2g14130 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C;Accession: E84514
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: E84514
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-789 <STO>
A;Cross-references: GB:AE002093; NID:g4587682; PIDN:AAD25852.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g14130
A;Map position: 2
C;Superfamily: Arabidopsis thaliana hypothetical protein T21L8.10

  Query Match             7.4%;  Score 99;  DB 2;  Length 789;
  Best Local Similarity   21.9%;  Pred. No. 15;
  Matches   59;  Conservative   38;  Mismatches   96;  Indels   76;  Gaps   12;

Qy      12 KEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNTVDLSSFDFRTGKMMPSKLQK 71
           :::| ||       :||    |: |: :|| :  : :||  |    |       |:  :|
Db     506 RKQVPRKRQKQVDSADVDV--PTRKEAKSKKR--KIIGNDGDNDDNDGDNDDFQPAPQRK 561

Qy      72 NKQRLRNDPLNQNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKVKSDPQRMNEQPRQ 131
           :|:             | ::|  |   ||   | |:     ||  ||  |  |: :
Db     562 SKRGT----------VPSIHTQAPF--TAEKKKHPIL----HPFAKV--DATRLEKLAE- 602

Qy     132 LFWEKRLQ----GLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSAVASALHTSSAPI 187
             |:|  :     :: : |  :   |:| |                          |
Db     603 --WKKSRKNKPLSIAGNIVDTKWFTTLETP-----------------------GKAIT 635

Qy     188 TGQVSAAVE-------------KNPAV------WLNTSQPLCKAFIVTDEDIRKQEERVQ 228
           |  | ||:|              || :|      :||        |:   |  : | | ::
Db     636 TTHVDAALELMKTRKESNPELFKNKSVVFVGSSFLNVIDESYMEFLDNKEGFQFQSEEIK 695

Qy     229 QVRKKLEEA---LMADILSRAADTEEMDI 254
           :|:||   |    |  |:      |:||:
Db     696 EVKKKYVAAYSIAMPYIVRNILKKEDMDV 724


RESULT 15
S49618
helicase-like transcription factor - human
C;Species: Homo sapiens (man)
C;Date: 05-Mar-1995 #sequence_revision 12-May-1995 #text_change 02-Aug-2002
C;Accession: S49618; A56095
R;Ding, H.; Descheemaeker, K.; Marynen, P.; Nelles, L.; Carvalho, T.; Carmo-Fonseca,
submitted to the EMBL Data Library, November 1994
A;Description: Characterization of a helicase-like transcription factor involved in p
A;Reference number: S49618
A;Accession: S49618
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1009 <DIN>
A;Cross-references: EMBL:Z46606; NID:g575250; PIDN:CAA86571.1; PID:g575251
R;Sheridan, P.L.; Schorpp, M.; Voz, M.L.; Jones, K.A.
J. Biol. Chem. 270, 4575-4587, 1995
A;Title: Cloning of an SNF2/SWI2-related protein that binds specifically to the SPH m
A;Reference number: A56095; MUID:95181452; PMID:7876228
A;Accession: A56095
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-34,'P',36-336,'D',338-381,'T',383-912,'K',914-1009 <SHE>
A;Cross-references: GB:L34673; NID:g531195; PIDN:AAA67436.1; PID:g531196
C;Genetics:
A;Gene: HLTF-1; GDB:SNF2L3; HIP116
A;Cross-references: GDB:392575
A;Map position: 3q25.1-3q26.1
C;Superfamily: human PML-1 protein; RING finger homology
C;Keywords: zinc
F;756-806/Domain: RING finger homology <RNG>

  Query Match             7.4%;  Score 99;  DB 2;  Length 1009;
  Best Local Similarity   22.3%;  Pred. No. 21;
  Matches   57;  Conservative   50;  Mismatches   79;  Indels   70;  Gaps   15;

Qy       6 LPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQ------LARYLGNTVDLSSF-- 57
```

```
          |||  |::            ::|:|| : :    : :|:      ||  :|    |::
Db    260 LPPFWEQ-----------RNDLYYNTITNFSEKDRPENVHGGILADDMGLGKTLTAIAV 307

Qy     58 ---DFRTGKMMPSKLQKNKQRLRNDPLNQNKGKPDL---NTTLPIRQTASIFKQPVTKVT 111
             :|  |: :|  ::: |: |     | | |   |   ||:    : |    :  :: :
Db    308 ILTNFHDGRPLP--IERVKKNLLKKEYNVNDGSMKLGGNNTS----EKADGLSKDASRCS 361

Qy    112 NHPS-NKVKSDPQ-RMNE----QPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGP 165
            || : :|   : ||:|    :|::   :  :| : :||  |  |:| |||: ::|
Db    362 EQPSISDIKEKSKFRMSELSSSRPKRR--KTAVQYIESSDSEE--IETSELPQKMKGKLK 417

Qy    166 GSNDET----------LLSAVASALH-TSSAPIT------------GQVSAAVEKNPAV 201
            ||          ::  || |   ||| | |              |     ||: |
Db    418 NVQSETKGRAKAGSSKVIEDVAFACALTSSVPTTKKKMLKKGACAVEGSKKTDVEERP-- 475

Qy    202 WLNTSQPLCKAFIVTD 217
            |:  :|   ::::
Db    476 --RTTLIICPLSVLSN 489
```

Search completed: March 12, 2003, 09:13:57
Job time : 26.0758 secs

GenCore version 5.1.4_p5_4578

OM protein - protein search, using sw model

Run on: March 12, 2003, 05:39:48 ; Search time 12.0684 Seconds
(without alignments)
915.501 Million cell updates/sec

Title: US-09-554-414B-2_COPY_150_411
Perfect score: 1344
Sequence: 1 MDCPALPPGWKKEEVIRKSG.....LSRAADTEEMDIEMDSGDEA 262

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 188354 seqs, 42170167 residues

Total number of hits satisfying chosen parameters: 188354

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1344	100.0	411	10	US-09-749-728B-1	Sequence 1, Appli
2	104	7.7	433	9	US-09-906-514-2	Sequence 2, Appli
3	100	7.4	2568	10	US-09-866-108-3	Sequence 3, Appli
4	91.5	6.8	608	10	US-09-799-777-7	Sequence 7, Appli
5	90.5	6.7	449	9	US-09-464-099A-5	Sequence 5, Appli
6	90.5	6.7	449	9	US-09-464-099A-7	Sequence 7, Appli
7	90.5	6.7	449	10	US-09-861-696-5	Sequence 5, Appli
8	90.5	6.7	449	10	US-09-861-696-7	Sequence 7, Appli
9	90.5	6.7	497	10	US-09-764-864-1314	Sequence 1314, Ap
10	90.5	6.7	534	10	US-09-764-864-861	Sequence 861, App
11	90.5	6.7	1111	10	US-09-815-242-12955	Sequence 12955, A
12	90	6.7	2665	10	US-09-864-761-34248	Sequence 34248, A
13	89	6.6	507	10	US-09-876-187-2	Sequence 2, Appli
14	89	6.6	507	10	US-09-749-728B-13	Sequence 13, Appl
15	88.5	6.6	428	9	US-09-906-514-4	Sequence 4, Appli
16	88.5	6.6	660	9	US-09-872-462-4	Sequence 4, Appli
17	88.5	6.6	939	10	US-09-226-248B-29	Sequence 29, Appl
18	88.5	6.6	939	10	US-09-801-368-380	Sequence 380, App
19	88	6.5	465	10	US-09-749-728B-17	Sequence 17, Appl
20	87	6.5	2285	10	US-09-932-183A-2	Sequence 2, Appli
21	86.5	6.4	459	10	US-09-988-915-1	Sequence 1, Appli
22	86	6.4	413	9	US-10-106-534-6	Sequence 6, Appli
23	86	6.4	478	9	US-09-925-299-903	Sequence 903, App
24	86	6.4	478	10	US-09-925-299-903	Sequence 903, App
25	86	6.4	26926	9	US-09-759-508B-2	Sequence 2, Appli
26	85.5	6.4	1404	10	US-09-811-045A-1	Sequence 1, Appli
27	84.5	6.3	229	10	US-09-988-915-7	Sequence 7, Appli
28	84.5	6.3	1043	10	US-09-946-805-4	Sequence 4, Appli
29	84	6.2	473	10	US-09-876-187-6	Sequence 6, Appli
30	84	6.2	473	10	US-09-876-187-8	Sequence 8, Appli
31	84	6.2	1093	10	US-09-801-368-392	Sequence 392, App
32	83	6.2	32	10	US-09-864-761-37960	Sequence 37960, A
33	83	6.2	96	10	US-09-864-761-36987	Sequence 36987, A
34	83	6.2	722	10	US-09-815-242-12888	Sequence 12888, A
35	83	6.2	792	9	US-10-025-380-1127	Sequence 1127, Ap
36	83	6.2	945	10	US-09-745-763-191	Sequence 191, App
37	83	6.2	991	10	US-09-815-242-5803	Sequence 5803, Ap
38	83	6.2	1711	10	US-09-771-161A-219	Sequence 219, App
39	83	6.2	1711	10	US-09-771-161A-220	Sequence 220, App
40	82.5	6.1	578	10	US-09-159-469-50	Sequence 50, Appl
41	82.5	6.1	578	10	US-09-798-042-50	Sequence 50, Appl
42	82.5	6.1	654	10	US-09-940-921B-4	Sequence 4, Appli
43	82	6.1	858	10	US-09-815-242-11396	Sequence 11396, A
44	82	6.1	1122	9	US-10-072-094-8	Sequence 8, Appli
45	82	6.1	1122	9	US-10-072-094-91	Sequence 91, Appl

ALIGNMENTS

```
RESULT 1
US-09-749-728B-1
; Sequence 1, Application US/09749728B
; Patent No. US20020142457A1
; GENERAL INFORMATION:
;  APPLICANT: Umezawa, Akihiro
;  APPLICANT:  Hata, Jun-Ichi
;  APPLICANT:  Fukuda, Keiichi
;  APPLICANT:  Ogawa, Satoshi
;  APPLICANT:  Sakurada, Kazuhiro
;  APPLICANT:  Gojo, Satoshi
;  APPLICANT:  Yamada, Yoji
;  TITLE OF INVENTION: THE CELL HAVING THE POTENTIALITY OF DIFFERENTIATION INTO CARDI
;  FILE REFERENCE: 00766.000043
;  CURRENT APPLICATION NUMBER: US/09/749,728B
;  CURRENT FILING DATE:  2001-09-17
;  PRIOR APPLICATION NUMBER: H11-372826
;  PRIOR FILING DATE: 1999-12-28
;  PRIOR APPLICATION NUMBER: PCT-JP00-01148
;  PRIOR FILING DATE: 2000-02-28
;  PRIOR APPLICATION NUMBER: PCT-JP00-07741
;  PRIOR FILING DATE: 2000-11-02
;  NUMBER OF SEQ ID NOS: 80
;  SOFTWARE: PatentIn Ver.2.0
; SEQ ID NO 1
;   LENGTH: 411
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-749-728B-1

  Query Match           100.0%;  Score 1344;  DB 10;  Length 411;
  Best Local Similarity  100.0%;  Pred. No. 2.7e-113;
  Matches  262;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MDCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNTVDLSSFDFR 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     150 MDCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNTVDLSSFDFR 209

Qy      61 TGKMMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKVKS 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     210 TGKMMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKVKS 269
```

```
Qy    121 DPQRMNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSAVASAL 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    270 DPQRMNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSAVASAL 329

Qy    181 HTSSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEALMA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    330 HTSSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEALMA 389

Qy    241 DILSRAADTEEMDIEMDSGDEA 262
          ||||||||||||||||||||||
Db    390 DILSRAADTEEMDIEMDSGDEA 411


RESULT 2
US-09-906-514-2
; Sequence 2, Application US/09906514
; Patent No. US20020170085A1
; GENERAL INFORMATION:
;  APPLICANT: Kaeppler, Shawn
;  APPLICANT:  Springer, Nathan
;  APPLICANT:  Phillips, Ronald
;  TITLE OF INVENTION: Methyl CpG Binding Domain Nucleic Acids  from Maize
;  FILE REFERENCE: Methybinding
;  CURRENT APPLICATION NUMBER: US/09/906,514
;  CURRENT FILING DATE:  2001-07-16
;  NUMBER OF SEQ ID NOS: 13
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 433
;   TYPE: PRT
;   ORGANISM: Zea mays
US-09-906-514-2

  Query Match             7.7%;  Score 104;  DB 9;  Length 433;
  Best Local Similarity   19.9%;  Pred. No. 0.15;
  Matches   53;  Conservative   48;  Mismatches   99;  Indels   66;  Gaps   11;

Qy      4 PALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYL-----GNTVDLSSFD 58
          || | || |    : : |  |:|:: : ||:|:: ::| ||::||     |  |  | ||
Db      2 PA-PDGWTK----KFTPLRGGRSEIVFVSPTGEEIKNKRQLSQYLKAHPGGPAV--SEFD 54

Qy     59 FRTGKMMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTASIFKQP----VTKVTNHP 114
          : ||               : |    :          | :   :|  |    : | : :
Db     55 WGTG--------------DTPRRSAR----------ISEKVKVFDSPEGEKIPKRSRNS 89

Qy    115 SNK-----VKSDPQRMNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSND 169
          | :      |  |:   |: :     |  :   : | ::    |:  : ::|    : |
Db     90 SGRKGKQGKKETPE--TEEAKDAETGKEAEEAPSEDAAKETDVEMKPAEEVKGASAETED 147

Qy    170 ETLLSAVASALHTSSAPITGQVSAAVEKNPAVW-----LNTSQPLCKAFIVTDEDIRKQE 224
          :  | | |        | ::: |:   |        :  ::|   |  |::
Db    148 ADMADAPAPAPMEEDKKQTEELAEAIAAPPVPSEEKKDVKPAEPEAAASNPTEDSAPAPA 207

Qy    225 ERVQQVRKKLEEALMADILSRAADTE 250
          |              ||: : ||:|:
Db    208 E------------PADVAAPAAETK 220


RESULT 3
US-09-866-108-3
; Sequence 3, Application US/09866108
; Patent No. US20020048800A1
; GENERAL INFORMATION:
;  APPLICANT: GU, Yizhong
;  APPLICANT:  JI, Yonggang
;  APPLICANT:  PENN, Sharron G.
;  APPLICANT:  HANZEL, David K.
;  APPLICANT:  RANK, David R.
;  APPLICANT:  CHEN, Wensheng
;  APPLICANT:  SHANNON, Mark
;  TITLE OF INVENTION: MYOSIN-LIKE GENE EXPRESSED IN HUMAN HEART AND MUSCLE
;  FILE REFERENCE: AEOMICA-7
;  CURRENT APPLICATION NUMBER: US/09/866,108
;  CURRENT FILING DATE:  2001-05-25
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 60/266,860
;  PRIOR FILING DATE: 2001-02-05
;  NUMBER OF SEQ ID NOS: 15752
;  SOFTWARE: Aeomica Sequence Listing Engine
; SEQ ID NO 3
;   LENGTH: 2568
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-866-108-3

  Query Match             7.4%;  Score 100;  DB 10;  Length 2568;
  Best Local Similarity   23.7%;  Pred. No. 4.1;
  Matches   42;  Conservative   24;  Mismatches   49;  Indels   62;  Gaps    7;

Qy    139 QGLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSAVASALHTSSAPITGQ-------- 190
          :|:      |::::|::| |  : |  | :   | : | | |      :  |
Db   1290 EGIDERKAVEELLETLDLEK--KAVAVGHSQVFLKAGVISRLEKQREKLVSQSIVLFQAA 1347

Qy    191 -----------------VSAAVEKNPAV---------W--LNTSQPLCKAFIVTDEDIR 221
                           :  ::|| ||          |  | : |||  | | | | :|
Db   1348 CKGFLSRQEFKKLKIRRLAAQCIQKNVAVFLAVKDWPWWQLLGSLQPLLSATIGT-EQLR 1406

Qy    222 KQEERVQQVRKKLE----------------EALMADILSRAADTEEMDIEMDSGDEA 262
           :|| :  :|:|||                |: :||: |  ||      |   || |
Db   1407 AKEEELTTLRRKLEKSEKLRNELRQNTDLLESKIADLTSDLAD------ERFKGDVA 1457


RESULT 4
US-09-799-777-7
; Sequence 7, Application US/09799777
; Patent No. US20020091244A1
;   GENERAL INFORMATION:
;        APPLICANT: Lal, Preeti
;                   Hillman, Jennifer L.
;                   Corley, Neil C.
;                   Guegler, Karl J.
;                   Baugh, Mariah
;                   Sather, Susan
;                   Shah, Purvi
;        TITLE OF INVENTION: HUMAN SIGNAL PEPTIDE-CONTAINING PROTEINS
;        NUMBER OF SEQUENCES: 154
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
;             STREET: 3174 PORTER DRIVE
```

```
;             CITY: PALO ALTO
;             STATE: CALIFORNIA
;             COUNTRY: USA
;             ZIP: 94304
;          COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
;          CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/799,777
;             FILING DATE: 06-Mar-2001
;             CLASSIFICATION: <Unknown>
;          PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/09/002,485
;             FILING DATE: <Unknown>
;          ATTORNEY/AGENT INFORMATION:
;             NAME: BILLINGS, LUCY J.
;             REGISTRATION NUMBER: 36,749
;             REFERENCE/DOCKET NUMBER: PF-0459 US
;          TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (650) 855-0555
;             TELEFAX: (650) 845-4166
;       INFORMATION FOR SEQ ID NO: 7:
;          SEQUENCE CHARACTERISTICS:
;             LENGTH: 608 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;          IMMEDIATE SOURCE:
;             LIBRARY: LUNGFET03
;             CLONE: 1255027
;          SEQUENCE DESCRIPTION: SEQ ID NO: 7 :
US-09-799-777-7

  Query Match             6.8%;  Score 91.5;  DB 10;  Length 608;
  Best Local Similarity   24.3%;  Pred. No. 3.2;
  Matches   57;  Conservative   30;  Mismatches   93;  Indels   55;  Gaps   10;

Qy      54 LSSFDFRTGKMMPSKL----------------QKNKQRLRNDPLNQNKGKPDLNTTLPIR 97
           |  |||    | : : :                :|| :| |:    |::|  |    :
Db      83 LQQFDFNVDKAVQAFVDGSAIQVLKEWNMTGKKKNNKRKRSKS-KQHQGNKDAKDKVERP 141

Qy      98 QTASIFKQPVTKVTNHPSN---KVKSDPQRMNEQPRQLFWEKRLQGLSASDVTEQIIKTM 154
           :   :  ||  :: | | |   |  |      ||:|  :  ||::  |
Db     142 EAGPLQPQP-PQIQNGPMNGCEKDSSSTDSANEKPALIPREKKISILE------------ 188

Qy     155 ELPKGLQGVGPGSNDETLLSAVASALHTSSAPITG--------QVSAAVEKNPA------ 200
           |  | |:||  |:    ||    | |  :  || |        | ||    ||:
Db     189 EPSKALRGVTEGNR---LLQQKLS-LDGNPKPIHGTTERSDGLQWSAEQPCNPSKPKAKT 244

Qy     201 --VWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEALMADILSRAADTEEMD 253
             |  ||  |     |  ||  :|:   :::  | |:    ::    |    ||:|
Db     245 SPVKSNT--PAAHLEIKPDELAKKRGPNIEKSVKDLQRCTVSLTRYRVMIKEEVD 297


RESULT 5
US-09-464-099A-5
; Sequence 5, Application US/09464099A
; Patent No. US20020168680A1
; GENERAL INFORMATION:
;  APPLICANT: Barry, Gerard F.
;  APPLICANT:  Kishore, Ganesh M.
;  APPLICANT:  Padgette, Stephen R.
;  APPLICANT:  Stallings, William C.
;  TITLE OF INVENTION: GLYPHOSATE TOLERANT 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASES
;  FILE REFERENCE: 11899.0175.CNUS01    MOBT:175-2
;  CURRENT APPLICATION NUMBER: US/09/464,099A
;  CURRENT FILING DATE:  1999-12-16
;  PRIOR APPLICATION NUMBER: US 09/137,440
;  PRIOR FILING DATE: 1998-08-20
;  PRIOR APPLICATION NUMBER: US 08/833,485
;  PRIOR FILING DATE: 1997-04-07
;  PRIOR APPLICATION NUMBER: US 08/306,063
;  PRIOR FILING DATE: 1994-09-13
;  PRIOR APPLICATION NUMBER: US 07/749,611
;  PRIOR FILING DATE: 1991-08-28
;  PRIOR APPLICATION NUMBER: US 07/576,537
;  PRIOR FILING DATE: 1990-08-31
;  NUMBER OF SEQ ID NOS: 70
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5
;   LENGTH: 449
;   TYPE: PRT
;   ORGANISM: Agrobacterium sp.
US-09-464-099A-5

  Query Match             6.7%;  Score 90.5;  DB 9;  Length 449;
  Best Local Similarity   22.0%;  Pred. No. 2.5;
  Matches   57;  Conservative   38;  Mismatches  123;  Indels   41;  Gaps    9;

Qy      16 IRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGN-------TVDL-SSFDFRTGKMMPS 67
           ||| |      ||:  :  |     :|: |   ||       |: |  ::| :|  :  :
Db      70 IRKEG------DVWIINGVGNGCLLQPEAALDFGNAGTGARLTMGLVGTYDMKTSFIGDA 123

Qy      68 KLQKNKQRLRNDPLNQNKGKPDL--NTTLPIRQTASIFKQPVTKVTNHPSNKVKSDPQRM 125
            | |       :|| :   : :      :|:         |:|      | :|||
Db     124 SLSKRPMGRVLNPLREMGVQVEAADGDRMPLTLIGPKTANPITYRVPMASAQVKS----- 178

Qy     126 NEQPRQLFWEKRLQGLSASDVT---EQIIKTMELPKGLQGVGPGSNDETLLSAVASALHT 182
                   | ||:   ||   | ::      | ||| |     ||    |     |
Db     179 ---------AVLLAGLNTPGVTTVIEPVMTRDHTEKMLQGFGADLTVETDKDGVRHIRIT 229

Qy     183 SSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEALMADI 242
           : || :  |  :|:   :|: ||  | :|   |:  :   :  |      |:  :
Db     230 GQGKLVGQ-TIDVPGDPS---STAFPLVAALLVEGSDVTIRNVLMNPTR----TGLILTL 281

Qy     243 LSRAADTEEMDIEMDSGDE 261
             || | ::  :  |::
Db     282 QEMGADIEVLNARLAGGED 300


RESULT 6
US-09-464-099A-7
; Sequence 7, Application US/09464099A
; Patent No. US20020168680A1
; GENERAL INFORMATION:
;  APPLICANT: Barry, Gerard F.
;  APPLICANT:  Kishore, Ganesh M.
;  APPLICANT:  Padgette, Stephen R.
;  APPLICANT:  Stallings, William C.
;  TITLE OF INVENTION: GLYPHOSATE TOLERANT 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHAS
;  FILE REFERENCE: 11899.0175.CNUS01    MOBT:175-2
;  CURRENT APPLICATION NUMBER: US/09/464,099A
;  CURRENT FILING DATE:  1999-12-16
;  PRIOR APPLICATION NUMBER: US 09/137,440
;  PRIOR FILING DATE: 1998-08-20
;  PRIOR APPLICATION NUMBER: US 08/833,485
;  PRIOR FILING DATE: 1997-04-07
;  PRIOR APPLICATION NUMBER: US 08/306,063
;  PRIOR FILING DATE: 1994-09-13
;  PRIOR APPLICATION NUMBER: US 07/749,611
;  PRIOR FILING DATE: 1991-08-28
;  PRIOR APPLICATION NUMBER: US 07/576,537
;  PRIOR FILING DATE: 1990-08-31
;  NUMBER OF SEQ ID NOS: 70
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7
;   LENGTH: 449
;   TYPE: PRT
;   ORGANISM: Pseudomonas sp.
US-09-464-099A-7

  Query Match             6.7%;  Score 90.5;  DB 9;  Length 449;
  Best Local Similarity   22.0%;  Pred. No. 2.5;
  Matches   57;  Conservative   38;  Mismatches  123;  Indels   41;  Gaps    9;
```

```
Qy      16 IRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGN-------TVDL-SSFDFRTGKMMPS 67
           ||| |      ||:  :  |     :|: |   ||       |: |  ::| :|  :  :
Db      70 IRKEG------DVWIINGVGNGCLLQPEAALDFGNAGTGARLTMGLVGTYDMKTSFIGDA 123

Qy      68 KLQKNKQRLRNDPLNQNKGKPDL--NTTLPIRQTASIFKQPVTKVTNHPSNKVKSDPQRM 125
            | |       :|| :   : :      :|:         |:|       | :|||
Db     124 SLSKRPMGRVLNPLREMGVQVEAADGDRMPLTLIGPKTANPITYRVPMASAQVKS----- 178

Qy     126 NEQPRQLFWEKRLQGLSASDVT---EQIIKTMELPKGLQGVGPGSNDETLLSAVASALHT 182
                     | ||:   ||   | ::      | ||| |     ||    |     |
Db     179 ---------AVLLAGLNTPGVTTVIEPVMTRDHTEKMLQGFGADLTVETDKDGVRHIRIT 229

Qy     183 SSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEALMADI 242
           : || :  |  :|:   :|: ||  | :|   |:  :   :  |      |:  :
Db     230 GQGKLVGQ-TIDVPGDPS---STAFPLVAALLVEGSDVTIRNVLMNPTR----TGLILTL 281

Qy     243 LSRAADTEEMDIEMDSGDE 261
               || | ::  :  |::
Db     282 QEMGADIEVLNARLAGGED 300
```

RESULT 7
US-09-861-696-5
; Sequence 5, Application US/09861696
; Patent No. US20020007053A1
; GENERAL INFORMATION:
; APPLICANT: Barry, Gerard F.
; APPLICANT: Kishore, Ganesh M.
; APPLICANT: Padgette, Stephen R.
; APPLICANT: Stallings, William C.
; TITLE OF INVENTION: GLYPHOSATE TOLERANT 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASES
; FILE REFERENCE: 11899.0175.CNUS04 MOBT:175-4
; CURRENT APPLICATION NUMBER: US/09/861,696
; CURRENT FILING DATE: 2001-05-21
; PRIOR APPLICATION NUMBER: US 09/137,440
; PRIOR FILING DATE: 1998-08-20
; PRIOR APPLICATION NUMBER: US 08/833,485
; PRIOR FILING DATE: 1997-04-07
; PRIOR APPLICATION NUMBER: US 08/306,063
; PRIOR FILING DATE: 1994-09-13
; PRIOR APPLICATION NUMBER: US 07/749,611
; PRIOR FILING DATE: 1991-08-28
; PRIOR APPLICATION NUMBER: US 07/576,537
; PRIOR FILING DATE: 1990-08-31
; NUMBER OF SEQ ID NOS: 70
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5
; LENGTH: 449
; TYPE: PRT
; ORGANISM: Agrobacterium sp.
US-09-861-696-5

Query Match 6.7%; Score 90.5; DB 10; Length 449;
Best Local Similarity 22.0%; Pred. No. 2.5;
Matches 57; Conservative 38; Mismatches 123; Indels 41; Gaps 9;

```
Qy      16 IRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGN-------TVDL-SSFDFRTGKMMPS 67
           ||| |      ||:  :  |     :|: |   ||       |: |  ::| :|  :  :
Db      70 IRKEG------DVWIINGVGNGCLLQPEAALDFGNAGTGARLTMGLVGTYDMKTSFIGDA 123

Qy      68 KLQKNKQRLRNDPLNQNKGKPDL--NTTLPIRQTASIFKQPVTKVTNHPSNKVKSDPQRM 125
            | |       :|| :   : :      :|:         |:|       | :|||
Db     124 SLSKRPMGRVLNPLREMGVQVEAADGDRMPLTLIGPKTANPITYRVPMASAQVKS----- 178

Qy     126 NEQPRQLFWEKRLQGLSASDVT---EQIIKTMELPKGLQGVGPGSNDETLLSAVASALHT 182
                     | ||:   ||   | ::      | ||| |     ||    |     |
Db     179 ---------AVLLAGLNTPGVTTVIEPVMTRDHTEKMLQGFGADLTVETDKDGVRHIRIT 229

Qy     183 SSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEALMADI 242
           : || :  |  :|:   :|: ||  | :|   |:  :   :  |      |:  :
Db     230 GQGKLVGQ-TIDVPGDPS---STAFPLVAALLVEGSDVTIRNVLMNPTR----TGLILTL 281

Qy     243 LSRAADTEEMDIEMDSGDE 261
               || | ::  :  |::
Db     282 QEMGADIEVLNARLAGGED 300
```

RESULT 8
US-09-861-696-7
; Sequence 7, Application US/09861696
; Patent No. US20020007053A1
; GENERAL INFORMATION:
; APPLICANT: Barry, Gerard F.
; APPLICANT: Kishore, Ganesh M.
; APPLICANT: Padgette, Stephen R.
; APPLICANT: Stallings, William C.
; TITLE OF INVENTION: GLYPHOSATE TOLERANT 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHAS
; FILE REFERENCE: 11899.0175.CNUS04 MOBT:175-4
; CURRENT APPLICATION NUMBER: US/09/861,696
; CURRENT FILING DATE: 2001-05-21
; PRIOR APPLICATION NUMBER: US 09/137,440
; PRIOR FILING DATE: 1998-08-20
; PRIOR APPLICATION NUMBER: US 08/833,485
; PRIOR FILING DATE: 1997-04-07
; PRIOR APPLICATION NUMBER: US 08/306,063
; PRIOR FILING DATE: 1994-09-13
; PRIOR APPLICATION NUMBER: US 07/749,611
; PRIOR FILING DATE: 1991-08-28
; PRIOR APPLICATION NUMBER: US 07/576,537
; PRIOR FILING DATE: 1990-08-31
; NUMBER OF SEQ ID NOS: 70
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7
; LENGTH: 449
; TYPE: PRT
; ORGANISM: Pseudomonas sp.
US-09-861-696-7

Query Match 6.7%; Score 90.5; DB 10; Length 449;
Best Local Similarity 22.0%; Pred. No. 2.5;
Matches 57; Conservative 38; Mismatches 123; Indels 41; Gaps 9;

```
Qy      16 IRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGN-------TVDL-SSFDFRTGKMMPS 67
           ||| |      ||:  :  |     :|: |   ||       |: |  ::| :|  :  :
Db      70 IRKEG------DVWIINGVGNGCLLQPEAALDFGNAGTGARLTMGLVGTYDMKTSFIGDA 123

Qy      68 KLQKNKQRLRNDPLNQNKGKPDL--NTTLPIRQTASIFKQPVTKVTNHPSNKVKSDPQRM 125
            | |       :|| :   : :      :|:         |:|       | :|||
Db     124 SLSKRPMGRVLNPLREMGVQVEAADGDRMPLTLIGPKTANPITYRVPMASAQVKS----- 178

Qy     126 NEQPRQLFWEKRLQGLSASDVT---EQIIKTMELPKGLQGVGPGSNDETLLSAVASALHT 182
                     | ||:   ||   | ::      | ||| |     ||    |     |
Db     179 ---------AVLLAGLNTPGVTTVIEPVMTRDHTEKMLQGFGADLTVETDKDGVRHIRIT 229

Qy     183 SSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEALMADI 242
           : || :  |  :|:   :|: ||  | :|   |:  :   :  |      |:  :
Db     230 GQGKLVGQ-TIDVPGDPS---STAFPLVAALLVEGSDVTIRNVLMNPTR----TGLILTL 281

Qy     243 LSRAADTEEMDIEMDSGDE 261
               || | ::  :  |::
Db     282 QEMGADIEVLNARLAGGED 300
```

RESULT 9
US-09-764-864-1314
; Sequence 1314, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ23
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792

```
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1314
;   LENGTH: 497
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (105)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-1314

  Query Match             6.7%;  Score 90.5;  DB 10;  Length 497;
  Best Local Similarity   19.5%;  Pred. No. 2.9;
  Matches   59;  Conservative   46;  Mismatches   97;  Indels  101;  Gaps   13;

Qy       7 PPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNTVDLSSFDFRTGKMMP 66
           |||| ||  :  ||                              || | |        |   |:
Db     109 PPGWLKEAELENSG----------------------LALYDGK-------DGTDGETEV 138

Qy      67 SKLQKNK------QRLRNDP---LNQNKGKPD---LNTTLPIR--QTASIFKQPVTKVTN 112
            ::|:||       :| | |   ::  :| ||   :  ::|::  |   :|   :|  :|
Db     139 GEIQQNKSVTYDLSKLVNYPGFNISTPRGIPDEWRIFGSIPMQACQQKDVFANYLT--SN 196

Qy     113 HPSNKVKSDPQRMNE-----QPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQ------ 161
             :  |||  :| :       |::   :|     : |    ::   ||:| | |
Db     197 FQAPGVKSGNKRSSSHSSPGSPKK---QKNESNSAGSPADMELDSDMEVPHGSQSSESFQ 253

Qy     162 -------------GVGPGSNDETLLSAVASALHTSSAPITGQVSAAVEKNPAVWLNTSQ 207
                       | |     |       |  | :| |   | ||:::
Db     254 FQPPLPPDTPPLPRGTPPPVFTPPLPKG-TPPLTPSDSPQTRTASGAVDED--------- 303

Qy     208 PLCKAFIVTDEDIRKQEERVQQVRKKLEEA-----------LMADILSRAADTEEMDIEM 256
                 :| |:: :|: |:    :: |             |  : :: :    |:|: :
Db     304 ------ALTLEELEEQQRRIWAALEQAESVNSDSDVPVDTPLTGNSVASSPCPNELDLPV 357

Qy     257 DSG 259
           |
Db     358 PEG 360


RESULT 10
US-09-764-864-861
; Sequence 861, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PTZ23
;  CURRENT APPLICATION NUMBER: US/09/764,864
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 1792
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 861
;   LENGTH: 534
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-764-864-861

  Query Match             6.7%;  Score 90.5;  DB 10;  Length 534;
  Best Local Similarity   19.5%;  Pred. No. 3.2;
  Matches   59;  Conservative   46;  Mismatches   97;  Indels  101;  Gaps   13;

Qy       7 PPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNTVDLSSFDFRTGKMMP 66
           |||| ||  :  ||                              || | |        |   |:
Db     146 PPGWLKEAELENSG----------------------LALYDGK-------DGTDGETEV 175

Qy      67 SKLQKNK------QRLRNDP---LNQNKGKPD---LNTTLPIR--QTASIFKQPVTKVTN 112
            ::|:||       :| | |   ::  :| ||   :  ::|::  |   :|   :|  :|
Db     176 GEIQQNKSVTYDLSKLVNYPGFNISTPRGIPDEWRIFGSIPMQACQQKDVFANYLT--SN 233

Qy     113 HPSNKVKSDPQRMNE-----QPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQ------ 161
             :  |||  :| :       |::   :|     : |    ::   ||:| | |
Db     234 FQAPGVKSGNKRSSSHSSPGSPKK---QKNESNSAGSPADMELDSDMEVPHGSQSSESFQ 290

Qy     162 -------------GVGPGSNDETLLSAVASALHTSSAPITGQVSAAVEKNPAVWLNTSQ 207
                       | |     |       |  | :| |   | ||:::
Db     291 FQPPLPPDTPPLPRGTPPPVFTPPLPKG-TPPLTPSDSPQTRTASGAVDED--------- 340

Qy     208 PLCKAFIVTDEDIRKQEERVQQVRKKLEEA-----------LMADILSRAADTEEMDIEM 256
                 :| |:: :|: |:    :: |             |  : :: :    |:|: :
Db     341 ------ALTLEELEEQQRRIWAALEQAESVNSDSDVPVDTPLTGNSVASSPCPNELDLPV 394

Qy     257 DSG 259
           |
Db     395 PEG 397


RESULT 11
US-09-815-242-12955
; Sequence 12955, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12955
;   LENGTH: 1111
;   TYPE: PRT
;   ORGANISM: Staphylococcus aureus
US-09-815-242-12955

  Query Match             6.7%;  Score 90.5;  DB 10;  Length 1111;
  Best Local Similarity   21.2%;  Pred. No. 9.1;
  Matches   55;  Conservative   34;  Mismatches   66;  Indels  105;  Gaps   12;

Qy      25 KSDVYYFSPSG-KKFRSKPQLARYLGN---------TVDLSSFDFRT--------GKMMP 66
           |   |  : :| :|   ||:      |          | |     |:|        | | |
Db     137 KPGTYVSAINGIEKETHKPKTHNMYSNNTNHRAKDSTPDYHKESFKTSEVPSAIFGTMKP 196

Qy      67 SKLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKVKSDPQRMN 126
            ||:                              : ||:|    || ||:|| |: :
Db     197 KKLENG------------------------------RIPVSK----PSEKVESDKQKYD 221

Qy     127 E--------QPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSAVAS 178
           :        | :||  ||:          : ::|        ||    |:|| : |
Db     222 KYVAKTQTSQNKQLEQEKQ---------NDSVVK--------QGTASKSSDENVSST--- 261

Qy     179 ALHTSSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEAL 238
           | | |   :|   ::          : :  : ||  |:||:: ||    ::: ::||
```

```
Db    262 ---TKSMPNYSKVDNTIK---------IENIYASQIV--EEIRRERERKVLQKRRFKKAL 307

Qy    239 M----------ADILSRAAD 248
                     | : || |
Db    308 QQKREEHKNEEQDAIQRAID 327


RESULT 12
US-09-864-761-34248
; Sequence 34248, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 34248
;   LENGTH: 2665
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AL034555.2
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL =  10
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 8.9
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.8
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL =  14
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 7.2
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 9.5
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 7.1
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 9.3
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 7.7
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL =  12
;   OTHER INFORMATION: EST_HUMAN HIT: AU117052.1, EVALUE 0.00e+00
;   OTHER INFORMATION: SWISSPROT HIT: P08640, EVALUE 3.00e-10
US-09-864-761-34248

  Query Match             6.7%;  Score 90;  DB 10;  Length 2665;
  Best Local Similarity   21.3%;  Pred. No. 34;
  Matches   65;  Conservative   27;  Mismatches   81;  Indels  132;  Gaps   14;

Qy      14 EVIRKSG-LSAGKSDVYYFSPSGKKFRSKPQLARYLGNTVDLSSFDFRTGKMMPSKLQKN 72
           ||: | | | | |                ||:      : |||             ||:
Db     333 EVVEKEGRLKARK-------------HLKPEQPADGVSAVDL------------EKLEAR 367

Qy      73 KQRLRNDPLNQNKGKPDLNTTLPIRQTASIF--------------------KQPVTK-V 110
           |:|  :  |   | ||::  : |  : | :                     ::|| | :
Db     368 KRRFADSNLKAEKQKPEVKKSSPEMEDARVLSKKQPDVSSREVILLREGEAERKPVRKEI 427

Qy     111 TNHPSNKVKSDPQRMNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQ-----GVGP 165
             | |:| |  |:|                          |:  ||  |      ||
Db     428 LKRESKKIKLD--RLN-------------------------TVASPKDCQELASISVGS 459

Qy     166 GSNDETLLSAVASALHTSSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTD------ED 219
           ||   : | |    |  |     :      | |      |:|  |   | |      ||
Db     460 GSRPSSDLQARLGELAGESV----ENQEVQSKKPI----PSKPQLKQLQVLDDQGPERED 511

Qy     220 IRKQ----------------------EERV-------QQVRKKLEEALMADILSRAADT 249
           :||                      ||::       |  ||::|:       ||
Db     512 VRKNYCSLRDETPERKSGQEKSHSVNTEEKIGIDIDHTQSYRKQMEQ-------SRRKQQ 564

Qy     250 EEMDI 254
             ||:|
Db     565 MEMEI 569


RESULT 13
US-09-876-187-2
; Sequence 2, Application US/09876187
; Patent No. US20020090603A1
; GENERAL INFORMATION:
;  APPLICANT: Lipton, Stuart A.
;  APPLICANT:  Okamoto, Shu-ichi
;  TITLE OF INVENTION: Methods of Differentiating and
;  TITLE OF INVENTION:  Protecting Cells By Modulating the P38/MEF2 Pathway
;  FILE REFERENCE: P-LJ 4714
;  CURRENT APPLICATION NUMBER: US/09/876,187
;  CURRENT FILING DATE:  2002-03-12
;  PRIOR APPLICATION NUMBER: US 60/209,539
;  PRIOR FILING DATE: 2000-06-05
;  NUMBER OF SEQ ID NOS: 23
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 507
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-876-187-2

  Query Match             6.6%;  Score 89;  DB 10;  Length 507;
  Best Local Similarity   26.7%;  Pred. No. 4.1;
  Matches   50;  Conservative   19;  Mismatches   50;  Indels   68;  Gaps   11;

Qy      32 SPSGKKF---RSKPQLARYLG-NTVDLSSFDFRTGKMMPSKLQKNKQRLRNDP------L 81
           || |  |   |: | |    | |::          ||:||:|         : |      |
Db     223 SPVGNGFVNSRASPNLIGATGANSL---------GKVMPTK---------SPPPPGGGNL 264

Qy      82 NQNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKVKSDPQRMNEQPRQLFWEKRLQGL 141
             |  ||||   :|                ||:|   |     |:       |   | :
Db     265 GMNSRKPDLRVVIP----------------PSSKGMMPPLSEEEE-----LELNTQRI 301

Qy     142 SASDVTEQI------IKTMELPKGLQGV------GPGSNDETLLSAVASALHTSSAP--- 186
           |:|  |: :      : |  ||   ||:         : | :| ||  |||   ::|
Db     302 SSSQATQPLATPVVSVTTPSLPP--QGLVYSAMPTAYNTDYSLTSADLSALQGFNSPGML 359
```

```
Qy     187 ITGQVSA 193
           |||||
Db     360 SLGQVSA 366


RESULT 14
US-09-749-728B-13
; Sequence 13, Application US/09749728B
; Patent No. US20020142457A1
; GENERAL INFORMATION:
;  APPLICANT: Umezawa, Akihiro
;  APPLICANT:  Hata, Jun-Ichi
;  APPLICANT:  Fukuda, Keiichi
;  APPLICANT:  Ogawa, Satoshi
;  APPLICANT:  Sakurada, Kazuhiro
;  APPLICANT:  Gojo, Satoshi
;  APPLICANT:  Yamada, Yoji
;  TITLE OF INVENTION: THE CELL HAVING THE POTENTIALITY OF DIFFERENTIATION INTO CARDIOMY
;  FILE REFERENCE: 00766.000043
;  CURRENT APPLICATION NUMBER: US/09/749,728B
;  CURRENT FILING DATE:  2001-09-17
;  PRIOR APPLICATION NUMBER: H11-372826
;  PRIOR FILING DATE: 1999-12-28
;  PRIOR APPLICATION NUMBER: PCT-JP00-01148
;  PRIOR FILING DATE: 2000-02-28
;  PRIOR APPLICATION NUMBER: PCT-JP00-07741
;  PRIOR FILING DATE: 2000-11-02
;  NUMBER OF SEQ ID NOS: 80
;  SOFTWARE: PatentIn Ver.2.0
; SEQ ID NO 13
;   LENGTH: 507
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-749-728B-13

  Query Match             6.6%;  Score 89;  DB 10;  Length 507;
  Best Local Similarity   26.7%;  Pred. No. 4.1;
  Matches   50;  Conservative   19;  Mismatches   50;  Indels   68;  Gaps   11;

Qy      32 SPSGKKF---RSKPQLARYLG-NTVDLSSFDFRTGKMMPSKLQKNKQRLRNDP------L 81
           || |  |   |: | |    | |::         ||:||:|         : |      |
Db     223 SPVGNGFVNSRASPNLIGATGANSL---------GKVMPTK---------SPPPPGGGNL 264

Qy      82 NQNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKVKSDPQRMNEQPRQLFWEKRLQGL 141
             |  ||||   :|                   ||:|    |    |:       |   | :
Db     265 GMNSRKPDLRVVIP-----------------PSSKGMMPPLSEEEE-----LELNTQRI 301

Qy     142 SASDVTEQI------IKTMELPKGLQGV------GPGSNDETLLSAVASALHTSSAP--- 186
           |:|  |: :      : |  ||   ||:          : | :| ||  |||   ::|
Db     302 SSSQATQPLATPVVSVTTPSLPP--QGLVYSAMPTAYNTDYSLTSADLSALQGFNSPGML 359

Qy     187 ITGQVSA 193
           |||||
Db     360 SLGQVSA 366


RESULT 15
US-09-906-514-4
; Sequence 4, Application US/09906514
; Patent No. US20020170085A1
; GENERAL INFORMATION:
;  APPLICANT: Kaeppler, Shawn
;  APPLICANT:  Springer, Nathan
;  APPLICANT:  Phillips, Ronald
;  TITLE OF INVENTION: Methyl CpG Binding Domain Nucleic Acids  from Maize
;  FILE REFERENCE: Methybinding
;  CURRENT APPLICATION NUMBER: US/09/906,514
;  CURRENT FILING DATE:  2001-07-16
;  NUMBER OF SEQ ID NOS: 13
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
;   LENGTH: 428
;   TYPE: PRT
;   ORGANISM: Zea mays
US-09-906-514-4

  Query Match             6.6%;  Score 88.5;  DB 9;  Length 428;
  Best Local Similarity   24.8%;  Pred. No. 3.6;
  Matches   37;  Conservative   28;  Mismatches   53;  Indels   31;  Gaps    8;

Qy       4 PALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYL----GNTVDLSSFDF 59
           || | || |:   ::      |:|:: : ||:|:: ::| ||::||    |     | ||:
Db       2 PA-PDGWTKKFTPQR----GGRSEIVFVSPTGEEIKNKRQLSQYLKAHPGGPA-ASDFDW 55

Qy      60 RTG-------------KMMPS-KLQKNKQRLRNDPLNQNK-GKPDLNTTLPIRQTASIFK 104
            ||             |:  | : :|  :| ||    : : || :   |   :   :
Db      56 GTGDTPRRSARISEKVKVFDSPEGEKIPKRSRNSSGRKGRQGKKEAPETEEAKDAETGQD 115

Qy     105 QP------VTKVTNHPSNKVKSDPQRMNE 127
           |       | |   |: : |  |   ::
Db     116 APSEDGTKETDVEMKPAEEAKEAPTETDD 144


Search completed: March 12, 2003, 09:14:42
Job time : 16.0684 secs
```

GenCore version 5.1.4_p5_4578

OM protein - protein search, using sw model

Run on: March 12, 2003, 00:59:01 ; Search time 38.1516 Seconds
(without alignments)
915.078 Million cell updates/sec

Title: US-09-554-414B-2_COPY_150_411
Perfect score: 1344
Sequence: 1 MDCPALPPGWKKEEVIRKSG.....LSRAADTEEMDIEMDSGDEA 262

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1344	100.0	263	19	AAW74980	Human secreted pro
2	1344	100.0	281	20	AAY73829	Human prostate tum
3	1344	100.0	281	20	AAY48439	Human prostate can
4	1344	100.0	411	20	AAY14197	Human DNA demethyl
5	1344	100.0	411	22	AAB99915	Human protein sequ
6	1344	100.0	411	22	AAG64314	Human protein #1.
7	1344	100.0	411	22	AAG64844	Heart muscle cell
8	1327	98.7	414	20	AAY14199	Mouse DNA demethyl
9	1050	78.1	282	23	AAE22578	Xenopus laevis MBD
10	1047.5	77.9	303	23	AAE22577	Xenopus laevis MBD
11	1021	76.0	291	20	AAY14198	Human DNA demethyl
12	1013	75.4	200	20	AAY07107	Colon cancer assoc
13	1012.5	75.3	285	20	AAY14200	Mouse DNA demethyl
14	753	56.0	223	22	AAB92997	Human protein sequ
15	739	55.0	218	22	AAB93733	Human protein sequ
16	436.5	32.5	314	22	ABB66035	Drosophila melanog
17	428	31.8	339	23	AAE22575	Drosophila dMBD-li
18	384.5	28.6	226	22	ABB63747	Drosophila melanog
19	384.5	28.6	226	23	AAE22576	Drosophila dMBD-li
20	341	25.4	64	23	AAE22569	Human MBD2 methyl
21	280.5	20.9	63	23	AAE22570	Xenopus laevis MBD
22	214.5	16.0	574	21	AAY28900	Regulator of fibro
23	214.5	16.0	590	21	AAB21008	Human nucleic acid
24	214.5	16.0	629	21	AAY28899	Regulator of fibro
25	214.5	16.0	630	21	AAY28898	Regulator of fibro
26	209	15.6	64	23	AAE22567	Human MBD1 methyl
27	166	12.4	219	21	AAG02051	Human secreted pro
28	144.5	10.8	68	23	AAE22566	Human MeCP2 methyl
29	139.5	10.4	51	23	AAE22574	Drosophila dMBD-li
30	126	9.4	580	20	AAW74473	Human MED1 endonuc
31	124.5	9.3	95	21	AAG50927	Arabidopsis thalia
32	124.5	9.3	149	21	AAG50926	Arabidopsis thalia
33	124.5	9.3	161	21	AAG50925	Arabidopsis thalia
34	123.5	9.2	95	21	AAG13952	Arabidopsis thalia
35	123.5	9.2	149	21	AAG13951	Arabidopsis thalia
36	123.5	9.2	161	21	AAG13950	Arabidopsis thalia
37	123.5	9.2	565	21	AAY44504	Human delta228-UV
38	120	8.9	68	23	AAE22568	Human MBD4 methyl
39	119	8.9	248	22	ABG17902	Novel human diagno
40	118	8.8	133	21	AAB24760	Plant SDF encoded
41	118	8.8	186	21	AAB24759	Plant SDF encoded
42	118	8.8	203	21	AAB24758	Plant SDF encoded
43	117.5	8.7	384	21	AAG42943	Arabidopsis thalia
44	117.5	8.7	384	21	AAG54911	Arabidopsis thalia
45	115	8.6	201	21	AAG38306	Arabidopsis thalia

ALIGNMENTS

RESULT 1
AAW74980
ID AAW74980 standard; Protein; 263 AA.
XX
AC AAW74980;
XX
DT 25-JAN-1999 (first entry)
XX
DE Human secreted protein encoded by gene 106 clone HT3AM65.
XX
KW Human; secreted protein; testis; tumour; foetal brain tissue;
KW fusion protein; cancer; central nervous system; seizure;
KW diagnosis; neurodegenerative disease.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 263
FT /label= unknown
XX
PN WO9839448-A2.
XX
PD 11-SEP-1998.
XX
PF 06-MAR-1998; 98WO-US04493.
XX
PR 02-OCT-1997; 97US-0061060.
PR 07-MAR-1997; 97US-0038621.
PR 07-MAR-1997; 97US-0040161.
PR 07-MAR-1997; 97US-0040162.
PR 07-MAR-1997; 97US-0040163.
PR 07-MAR-1997; 97US-0040333.
PR 07-MAR-1997; 97US-0040334.

PR 07-MAR-1997; 97US-0040336.
PR 07-MAR-1997; 97US-0040626.
PR 11-APR-1997; 97US-0043311.
PR 11-APR-1997; 97US-0043312.
PR 11-APR-1997; 97US-0043313.
PR 11-APR-1997; 97US-0043314.
PR 11-APR-1997; 97US-0043568.
PR 11-APR-1997; 97US-0043569.
PR 11-APR-1997; 97US-0043576.
PR 11-APR-1997; 97US-0043578.
PR 11-APR-1997; 97US-0043580.
PR 11-APR-1997; 97US-0043669.
PR 11-APR-1997; 97US-0043670.
PR 11-APR-1997; 97US-0043671.
PR 11-APR-1997; 97US-0043672.
PR 11-APR-1997; 97US-0043674.
PR 23-MAY-1997; 97US-0047492.
PR 23-MAY-1997; 97US-0047500.
PR 23-MAY-1997; 97US-0047501.
PR 23-MAY-1997; 97US-0047502.
PR 23-MAY-1997; 97US-0047503.
PR 23-MAY-1997; 97US-0047581.
PR 23-MAY-1997; 97US-0047582.
PR 23-MAY-1997; 97US-0047583.
PR 23-MAY-1997; 97US-0047584.
PR 23-MAY-1997; 97US-0047585.
PR 23-MAY-1997; 97US-0047586.
PR 23-MAY-1997; 97US-0047587.
PR 23-MAY-1997; 97US-0047588.
PR 23-MAY-1997; 97US-0047589.
PR 23-MAY-1997; 97US-0047590.
PR 23-MAY-1997; 97US-0047592.
PR 23-MAY-1997; 97US-0047593.
PR 23-MAY-1997; 97US-0047594.
PR 23-MAY-1997; 97US-0047595.
PR 23-MAY-1997; 97US-0047596.
PR 23-MAY-1997; 97US-0047597.
PR 23-MAY-1997; 97US-0047598.
PR 23-MAY-1997; 97US-0047599.
PR 23-MAY-1997; 97US-0047600.
PR 23-MAY-1997; 97US-0047601.
PR 23-MAY-1997; 97US-0047612.
PR 23-MAY-1997; 97US-0047613.
PR 23-MAY-1997; 97US-0047614.
PR 23-MAY-1997; 97US-0047615.
PR 23-MAY-1997; 97US-0047617.
PR 23-MAY-1997; 97US-0047618.
PR 23-MAY-1997; 97US-0047632.
PR 23-MAY-1997; 97US-0047633.
PR 06-JUN-1997; 97US-0048964.
PR 06-JUN-1997; 97US-0048974.
PR 13-JUN-1997; 97US-0049610.
PR 08-JUL-1997; 97US-0051926.
PR 16-JUL-1997; 97US-0052874.
PR 18-AUG-1997; 97US-0055724.
PR 22-AUG-1997; 97US-0056630.
PR 22-AUG-1997; 97US-0056631.
PR 22-AUG-1997; 97US-0056632.
PR 22-AUG-1997; 97US-0056636.
PR 22-AUG-1997; 97US-0056637.
PR 22-AUG-1997; 97US-0056662.
PR 22-AUG-1997; 97US-0056664.
PR 22-AUG-1997; 97US-0056845.
PR 22-AUG-1997; 97US-0056862.
PR 22-AUG-1997; 97US-0056864.
PR 22-AUG-1997; 97US-0056872.
PR 22-AUG-1997; 97US-0056874.
PR 22-AUG-1997; 97US-0056875.
PR 22-AUG-1997; 97US-0056876.
PR 22-AUG-1997; 97US-0056877.
PR 22-AUG-1997; 97US-0056878.
PR 22-AUG-1997; 97US-0056879.
PR 22-AUG-1997; 97US-0056880.
PR 22-AUG-1997; 97US-0056881.
PR 22-AUG-1997; 97US-0056882.
PR 22-AUG-1997; 97US-0056884.
PR 22-AUG-1997; 97US-0056886.
PR 22-AUG-1997; 97US-0056887.
PR 22-AUG-1997; 97US-0056888.
PR 22-AUG-1997; 97US-0056889.
PR 22-AUG-1997; 97US-0056892.
PR 22-AUG-1997; 97US-0056893.
PR 22-AUG-1997; 97US-0056894.
PR 22-AUG-1997; 97US-0056903.
PR 22-AUG-1997; 97US-0056908.
PR 22-AUG-1997; 97US-0056909.
PR 22-AUG-1997; 97US-0056910.
PR 22-AUG-1997; 97US-0056911.
PR 05-SEP-1997; 97US-0057650.
PR 05-SEP-1997; 97US-0057669.
PR 05-SEP-1997; 97US-0057761.
PR 12-SEP-1997; 97US-0058785.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Bednarik DP, Brewer LA, Carter KC, Duan R, Ebner R, Endress GA;
PI Feng P, Ferrie AM, Fischer CL, Florence KA, Greene JM, Hu JS;
PI Kyaw H, Lafleur DW, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA;
PI Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;
XX
DR WPI; 1998-506364/43.
DR N-PSDB; AAV59765.
XX
PT New isolated human genes and the secreted polypeptide(s) they encode
PT - useful for diagnosis and treatment of e.g. cancers, neurological
PT disorders, immune diseases, inflammation or blood disorders
XX
PS Claim 1; Page 686-687; 721pp; English.
XX
CC This sequence represents a secreted human protein encoded by the nucleic
CC acid molecule designated Gene 106 from the human cDNA clone HT3AM65
CC (deposited as clone ATCC 97901 and ATCC 209047).
CC The gene can be used to generate fusion proteins by linking to the gene
CC to a human immunoglobulin Fc portion (e.g. AAV59502) for increasing the
CC stability of the fused protein as compared to the human protein only.
CC The invention relates to 186 novel genes and their fragments (nucleic
CC acid sequences: AAV59511-V59812; amino acid sequences AAW74731-W75026)
CC which are useful for preventing, treating or ameliorating medical
CC conditions e.g. by protein or gene therapy. Also, pathological
CC conditions can be diagnosed by determining the amount of the new
CC polypeptides in a sample or by determining the presence of mutations in
CC the new polynucleotides. Specific uses are described for each of the 186
CC polynucleotides, based on which tissues they are most highly expressed in
CC (see AAV59511 for described uses).
XX
SQ Sequence 263 AA;

Query Match 100.0%; Score 1344; DB 19; Length 263;
Best Local Similarity 100.0%; Pred. No. 1.5e-119;
Matches 262; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 MDCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNTVDLSSFDFR 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MDCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNTVDLSSFDFR 60

Qy      61 TGKMMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKVKS 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TGKMMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKVKS 120

Qy     121 DPQRMNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSAVASAL 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 DPQRMNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSAVASAL 180

Qy     181 HTSSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEALMA 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 HTSSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEALMA 240
```

```
Qy    241 DILSRAADTEEMDIEMDSGDEA 262
          |||||||||||||||||||||
Db    241 DILSRAADTEEMDIEMDSGDEA 262
```

RESULT 2
AAY73829
ID AAY73829 standard; Protein; 281 AA.
XX
AC AAY73829;
XX
DT 14-MAR-2000 (first entry)
XX
DE Human prostate tumor EST fragment derived protein #16.
XX
KW Pancreas; tumor; EST; expressed sequence tag; human; cytostatic;
KW treatment.
XX
OS Homo sapiens.
XX
PN DE19820190-A1.
XX
PD 04-NOV-1999.
XX
PF 28-APR-1998; 98DE-1020190.
XX
PR 28-APR-1998; 98DE-1020190.
XX
PA (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;
XX
DR WPI; 1999-621386/54.
DR N-PSDB; AAZ52863.
XX
PT New human nucleic acid sequences from pancreatic tumors, and related
PT proteins -
XX
PS Claim 23; Page 315; 502pp; German.
XX
CC This invention describes novel polypeptides and their encoding nucleic
CC acids derived from human pancreatic tumor tissue which have cytostatic
CC activity. The sequences are also useful in producing pharmaceutical
CC compositions for treatment of pancreatic tumors. AAY73814-Y74252
CC represent protein fragments encoded by the human pancreatic tumor cDNA
CC library derived expressed sequence tag (EST) sequences represented in
CC AAZ52858-Z53014.
XX
SQ Sequence 281 AA;

```
  Query Match            100.0%;  Score 1344;  DB 20;  Length 281;
  Best Local Similarity  100.0%;  Pred. No. 1.6e-119;
  Matches  262;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 MDCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNTVDLSSFDFR 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     20 MDCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNTVDLSSFDFR 79

Qy     61 TGKMMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKVKS 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     80 TGKMMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKVKS 139

Qy    121 DPQRMNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSAVASAL 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    140 DPQRMNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSAVASAL 199

Qy    181 HTSSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEALMA 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    200 HTSSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEALMA 259

Qy    241 DILSRAADTEEMDIEMDSGDEA 262
          ||||||||||||||||||||||
```

```
Db    260 DILSRAADTEEMDIEMDSGDEA 281
```

RESULT 3
AAY48439
ID AAY48439 standard; Protein; 281 AA.
XX
AC AAY48439;
XX
DT 08-DEC-1999 (first entry)
XX
DE Human prostate cancer-associated protein 136.
XX
KW Expressed sequence tag; EST; prostate; tumor; treatment; gene therapy;
KW cancer; tissue specificity; human.
XX
OS Homo sapiens.
XX
PN DE19811194-A1.
XX
PD 16-SEP-1999.
XX
PF 10-MAR-1998; 98DE-1011194.
XX
PR 10-MAR-1998; 98DE-1011194.
XX
PA (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
PI Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E, Rosenthal A;
XX
DR WPI; 1999-519629/44.
DR N-PSDB; AAZ33533.
XX
PT New nucleic acid expressed at high level in normal prostatic tissue and
PT encoded polypeptides, used to treat cancer and screen for therapeutic
PT agents -
XX
PS Claim 22; 176; 194pp; German.
XX
CC This invention describes novel nucleic acid sequences (A) that are
CC expressed at high level in normal prostatic tissue. Polypeptides (I)
CC encoded by (A) are used: (a) for identifying agents for treatment of
CC prostatic cancer and (b) for therapy of prostate cancer, optionally
CC where expressed by gene therapy methods. (A) is also used to isolate
CC full-length genes (for gene therapy) and for recombinant production of
CC (I), which can be used to raise specific antibodies. (A) are identified
CC by assembly of ESTs (expressed sequence tags) before these are analyzed
CC for expression pattern (tissue specificity). This approach eliminates
CC many of the false results, as regards tissue specificity, associated
CC with known methods that use single (usually short) ESTs. AAY48304-Y48456
CC represent peptides encoded by the expressed sequence tags described in
CC the method of the invention.
XX
SQ Sequence 281 AA;

```
  Query Match            100.0%;  Score 1344;  DB 20;  Length 281;
  Best Local Similarity  100.0%;  Pred. No. 1.6e-119;
  Matches  262;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 MDCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNTVDLSSFDFR 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     20 MDCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNTVDLSSFDFR 79

Qy     61 TGKMMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKVKS 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     80 TGKMMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKVKS 139

Qy    121 DPQRMNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSAVASAL 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    140 DPQRMNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSAVASAL 199

Qy    181 HTSSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEALMA 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db     200 HTSSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEALMA 259

Qy     241 DILSRAADTEEMDIEMDSGDEA 262
           |||||||||||||||||||||
Db     260 DILSRAADTEEMDIEMDSGDEA 281


RESULT 4
AAY14197
ID   AAY14197 standard; Protein; 411 AA.
XX
AC   AAY14197;
XX
DT   28-JUL-1999  (first entry)
XX
DE   Human DNA demethylase, dMTase1, protien sequence.
XX
KW   DNA demethylase; dMTase1; dMTase2; human; cancer cell; beta-thalassemia;
KW   DNA methylation pattern; methylation inhibitor; gene therapy; diagnosis;
KW   genetic defect correction; sickle cell anaemia.
XX
OS   Homo sapiens.
XX
PN   WO9924583-A1.
XX
PD   20-MAY-1999.
XX
PF   12-NOV-1998;   98WO-CA01059.
XX
PR   11-MAY-1998;   98CA-2230991.
PR   12-NOV-1997;   97CA-2220805.
XX
PA   (UYMC-) UNIV MCGILL.
XX
PI   Bhattacharya S,  Ramchandani S,  Szyf M;
XX
DR   WPI; 1999-347283/29.
DR   N-PSDB; AAX61218.
XX
PT   Human and murine DNA demethylases useful in the diagnosis of cancer
XX
PS   Disclosure; Fig 9c; 114pp; English.
XX
CC   This sequence is the human DNA demethylase, designated dMTase1, of
CC   the invention. The DNA demethylase is overexpressed in cancer cells.
CC   Expression of the demethylase cDNA is useful to alter DNA methylation
CC   patterns of DNA in vitro in cells or in vivo in humans, animals and
CC   plants. The cDNA is in antisense orientation to inhibit demethylase in
CC   cancer cells for therapeutic purposes. The demethylase is used to alter
CC   the differentiation state and to generate stem cells for therapeutics,
CC   cells for animal cloning and to improve expression of foreign genes. The
CC   cDNA can also be used for recombinant production of large amounts of the
CC   demethylase. The protein can be used to raise antibodies against
CC   demethylase. It can also be used for high throughput screening of
CC   demethylase inhibitors, and for obtaining the x-ray crystal structure.
CC   The demethylase cDNA and protein are also useful for changing the state
CC   of differentiation of a cell to allow gene therapy, stem cell selection
CC   or cell cloning; or for inhibiting methylation in cancer cells using
CC   vector mediated gene therapy. Antagonists or inhibitors of the
CC   demethylase can be used to manufacture medicaments for cancer treatment,
CC   for restoring an aberrant methylation patterns or changing methylation
CC   patterns in patient DNA. Change of the methylation pattern activates a
CC   silent gene permitting the correction of a genetic defect, such as
CC   beta-thalassemia or sickle cell anaemia. The cDNA can be used as a
CC   template to design antisense oligonucleotides and ribozymes. The cDNA can
CC   also be used in two-hybrid systems in yeast to identify proteins
CC   interacting with demethylase. Determining the level of DNA methylase
CC   expression in a cell can be used as a method for diagnosis of cancer,
CC   where overexpression is indicative of cancer cells.
XX
SQ   Sequence   411 AA;

  Query Match          100.0%;  Score 1344;  DB 20;  Length 411;
  Best Local Similarity   100.0%;  Pred. No. 2.8e-119;
  Matches  262;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MDCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNTVDLSSFDFR 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     150 MDCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNTVDLSSFDFR 209

Qy      61 TGKMMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKVKS 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     210 TGKMMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKVKS 269

Qy     121 DPQRMNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSAVASAL 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     270 DPQRMNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSAVASAL 329

Qy     181 HTSSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEALMA 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     330 HTSSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEALMA 389

Qy     241 DILSRAADTEEMDIEMDSGDEA 262
           ||||||||||||||||||||||
Db     390 DILSRAADTEEMDIEMDSGDEA 411


RESULT 5
AAB99915
ID   AAB99915 standard; Protein; 411 AA.
XX
AC   AAB99915;
XX
DT   26-SEP-2001  (first entry)
XX
DE   Human protein sequence SEQ ID NO:1.
XX
KW   Differentiation; heart muscle cell; cytokine; transcription factor;
KW   proliferation; surface antigen; heart disease; cardiomyocyte;
KW   bone marrow; umbilical blood cell; heart muscle degeneration;
KW   myocardial infarction.
XX
OS   Homo sapiens.
XX
PN   WO200148150-A1.
XX
PD   05-JUL-2001.
XX
PF   02-NOV-2000; 2000WO-JP07741.
XX
PR   28-DEC-1999;   99JP-0372826.
PR   28-FEB-2000; 2000WO-JP01148.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI   Umezawa A,  Hata J,  Fukuda K,  Ogawa S,  Sakurada K,  Gojo S;
PI   Yamada Y;
XX
DR   WPI; 2001-425655/45.
DR   N-PSDB; AAH44351.
XX
PT   Cells capable of differentiating into cardiomyocytes and originating in
PT   bone marrow or umbilical blood cells for study of cardiomyocyte
PT   differentiation and treatment of heart disease -
XX
PS   Claim 22; Page 84-86; 187pp; Japanese.
XX
CC   The present invention describes cells originating in bone marrow or
CC   umbilical blood cells which are capable of differentiating into
CC   cardiomyocytes. Also described are: (1) cardiomyocytes produced by the
CC   differentiation of the cells; (2) a method for carrying out the
CC   differentiation into cardiomyocytes, regulated by a promotional and/or
CC   inhibitory factor; (3) a method for the differentiation of the cells
CC   into cell types other than cardiomyocytes; (4) drug compositions
CC   promoting the formation of heart muscle and regeneration of heart tissue
CC   which contain the cells; (5) a method for the production of antibodies
```

CC which recognise the cells, especially antibodies which recognise a
CC surface antigen on the cells; (6) a method for screening factors which
CC promote the proliferation of the cells; (7) a method for immortalising
CC the cells by expressing telomerase in them; (8) drug compositions for
CC the treatment of heart disease which contain the immortalised cells; and
CC (9) cell-free supernatant from the culture of the cells and its use in
CC promoting their differentiation into cardiomyocytes. The cells are used
CC in the treatment of diseases involving heart muscle degeneration, such
CC as myocardial infarction and in the study of cardiomyocyte
CC differentiation. AAH44351 to AAH44409 and AAB99915 to AAB99935 represent
CC sequences used in the exemplification of the present invention.
XX
SQ Sequence 411 AA;

Query Match 100.0%; Score 1344; DB 22; Length 411;
Best Local Similarity 100.0%; Pred. No. 2.8e-119;
Matches 262; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 MDCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNTVDLSSFDFR 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     150 MDCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNTVDLSSFDFR 209

Qy      61 TGKMMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKVKS 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     210 TGKMMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKVKS 269

Qy     121 DPQRMNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSAVASAL 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     270 DPQRMNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSAVASAL 329

Qy     181 HTSSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEALMA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     330 HTSSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEALMA 389

Qy     241 DILSRAADTEEMDIEMDSGDEA 262
          ||||||||||||||||||||||
Db     390 DILSRAADTEEMDIEMDSGDEA 411
```

RESULT 6
AAG64314
ID AAG64314 standard; Protein; 411 AA.
XX
AC AAG64314;
XX
DT 24-SEP-2001 (first entry)
XX
DE Human protein #1.
XX
KW Angiogenesis; cardiant; cell differentiating agent; bone marrow;
KW heart muscle cell; heart disease; human.
XX
OS Homo sapiens.
XX
PN WO200148149-A1.
XX
PD 05-JUL-2001.
XX
PF 28-FEB-2000; 2000WO-JP01148.
XX
PR 28-DEC-1999; 99JP-0372826.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI Umezawa A, Hata J, Fukuda K, Ogawa S, Sakurada K;
XX
DR WPI; 2001-418252/44.
DR N-PSDB; AAH49586.
XX
PT New adult bone marrow-originated cells capable of differentiating into
PT heart muscle cells, applicable as remedies for various heart diseases
PT particularly with damaged heart muscle accompanying degeneration -
XX
PS Claim 12; Pages 55-57; 158pp; Japanese.
XX
CC The present invention relates to cells isolated from bone marrow, which
CC are capable of at least differentiating into heart muscle cells. The
CC cells are applicable as remedies for various heart diseases particularly
CC with damaged heart muscle accompanying degeneration. The present sequence
CC was used to illustrate the present invention.
XX
SQ Sequence 411 AA;

Query Match 100.0%; Score 1344; DB 22; Length 411;
Best Local Similarity 100.0%; Pred. No. 2.8e-119;
Matches 262; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 MDCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNTVDLSSFDFR 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     150 MDCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNTVDLSSFDFR 209

Qy      61 TGKMMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKVKS 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     210 TGKMMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKVKS 269

Qy     121 DPQRMNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSAVASAL 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     270 DPQRMNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSAVASAL 329

Qy     181 HTSSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEALMA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     330 HTSSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEALMA 389

Qy     241 DILSRAADTEEMDIEMDSGDEA 262
          ||||||||||||||||||||||
Db     390 DILSRAADTEEMDIEMDSGDEA 411
```

RESULT 7
AAG64844
ID AAG64844 standard; Protein; 411 AA.
XX
AC AAG64844;
XX
DT 21-SEP-2001 (first entry)
XX
DE Heart muscle cell differentiation related protein SEQ ID NO: 1.
XX
KW Heart muscle cell; human; cell differentiation; heart disease.
XX
OS Homo sapiens.
XX
PN WO200148151-A1.
XX
PD 05-JUL-2001.
XX
PF 27-DEC-2000; 2000WO-JP09323.
XX
PR 28-DEC-1999; 99JP-0372826.
PR 28-FEB-2000; 2000WO-JP01148.
PR 02-NOV-2000; 2000WO-JP07741.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI Umezawa A, Hata J, Fukuda K, Ogawa S, Sakurada K, Gojo S;
PI Yamada Y;
XX
DR WPI; 2001-425656/45.
DR N-PSDB; AAH48220.
XX
PT Cells capable of differentiating into cardiomyocytes and originating in
PT bone marrow or umbilical blood cells for study of cardiomyocyte
PT differentiation and treatment of heart disease -
XX
PS Claim 28; Page 90-92; 183pp; Japanese.
XX

```
CC   The present invention provides cells originating in the human bone marrow
CC   or umbilical blood cells which are capable of differentiating into
CC   cardiomyocytes. These cells are useful in the treatment of diseases
CC   involving heart muscle degeneration, such as myocardial infarction, and
CC   the study of cardiomyocyte differentiation. The present sequence is
CC   a protein described in the exemplification of the invention.
XX
SQ   Sequence   411 AA;

  Query Match             100.0%;  Score 1344;  DB 22;  Length 411;
  Best Local Similarity   100.0%;  Pred. No. 2.8e-119;
  Matches  262;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MDCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNTVDLSSFDFR 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     150 MDCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNTVDLSSFDFR 209

Qy      61 TGKMMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKVKS 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     210 TGKMMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKVKS 269

Qy     121 DPQRMNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSAVASAL 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     270 DPQRMNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSAVASAL 329

Qy     181 HTSSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEALMA 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     330 HTSSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEALMA 389

Qy     241 DILSRAADTEEMDIEMDSGDEA 262
           ||||||||||||||||||||||
Db     390 DILSRAADTEEMDIEMDSGDEA 411


RESULT 8
AAY14199
ID   AAY14199 standard; Protein; 414 AA.
XX
AC   AAY14199;
XX
DT   28-JUL-1999  (first entry)
XX
DE   Mouse DNA demethylase, dMTase1, protien sequence.
XX
KW   DNA demethylase; dMTase1; dMTase2; mouse; cancer cell; beta-thalassemia;
KW   DNA methylation pattern; methylation inhibitor; gene therapy; diagnosis;
KW   genetic defect correction; sickle cell anaemia.
XX
OS   Mus sp.
XX
PN   WO9924583-A1.
XX
PD   20-MAY-1999.
XX
PF   12-NOV-1998;   98WO-CA01059.
XX
PR   11-MAY-1998;   98CA-2230991.
PR   12-NOV-1997;   97CA-2220805.
XX
PA   (UYMC-) UNIV MCGILL.
XX
PI   Bhattacharya S,  Ramchandani S,  Szyf M;
XX
DR   WPI; 1999-347283/29.
DR   N-PSDB; AAX61220.
XX
PT   Human and murine DNA demethylases useful in the diagnosis of cancer
XX
PS   Disclosure; Fig 9k; 114pp; English.
XX
CC   This sequence is the mouse DNA demethylase, designated dMTase1, of
CC   the invention. The DNA demethylase is overexpressed in cancer cells.
CC   Expression of the demethylase cDNA is useful to alter DNA methylation
CC   patterns of DNA in vitro in cells or in vivo in humans, animals and
CC   plants. The cDNA is in antisense orientation to inhibit demethylase in
CC   cancer cells for therapeutic purposes. The demethylase is used to alter
CC   the differentiation state and to generate stem cells for therapeutics,
CC   cells for animal cloning and to improve expression of foreign genes. The
CC   cDNA can also be used for recombinant production of large amounts of the
CC   demethylase. The protein can be used to raise antibodies against
CC   demethylase. It can also be used for high throughput screening of
CC   demethylase inhibitors, and for obtaining the x-ray crystal structure.
CC   The demethylase cDNA and protein are also useful for changing the state
CC   of differentiation of a cell to allow gene therapy, stem cell selection
CC   or cell cloning; or for inhibiting methylation in cancer cells using
CC   vector mediated gene therapy. Antagonists or inhibitors of the
CC   demethylase can be used to manufacture medicaments for cancer treatment,
CC   for restoring an aberrant methylation patterns or changing methylation
CC   patterns in patient DNA. Change of the methylation pattern activates a
CC   silent gene permitting the correction of a genetic defect, such as
CC   beta-thalassemia or sickle cell anaemia. The cDNA can be used as a
CC   template to design antisense oligonucleotides and ribozymes. The cDNA can
CC   also be used in two-hybrid systems in yeast to identify proteins
CC   interacting with demethylase. Determining the level of DNA methylase
CC   expression in a cell can be used as a method for diagnosis of cancer,
CC   where overexpression is indicative of cancer cells.
XX
SQ   Sequence   414 AA;

  Query Match             98.7%;  Score 1327;  DB 20;  Length 414;
  Best Local Similarity   98.5%;  Pred. No. 1.2e-117;
  Matches  258;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 MDCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNTVDLSSFDFR 60
           ||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db     153 MDCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNAVDLSSFDFR 212

Qy      61 TGKMMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKVKS 120
           ||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db     213 TGKMMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTASIFKQPVTKFTNHPSNKVKS 272

Qy     121 DPQRMNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSAVASAL 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     273 DPQRMNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSAVASAL 332

Qy     181 HTSSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEALMA 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     333 HTSSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEALMA 392

Qy     241 DILSRAADTEEMDIEMDSGDEA 262
           ||||||||||:||:|||||||
Db     393 DILSRAADTEEVDIDMDSGDEA 414


RESULT 9
AAE22578
ID   AAE22578 standard; Protein; 282 AA.
XX
AC   AAE22578;
XX
DT   26-JUL-2002  (first entry)
XX
DE   Xenopus laevis MBD3 protein.
XX
KW   Gene expression; cellular chromatin; methyl CpG binding domain; cancer;
KW   localisation domain; diabetic retinopathy; ischaemia; HIV infection;
KW   human immuno deficiency virus; macular degeneration; vascular disease;
KW   rheumatoid arthritis; psoriasis; Alzheimer's disease; muscular dystrophy;
KW   sickle cell anaemia; stroke; neurodegenerative disease; cystic fibrosis;
KW   gene therapy; cytostatic; antidiabetic; ophthalmological; vasotropic;
KW   neuroprotective; nootropic; cerebroprotective; antibacterial; antifungal;
KW   antiviral; MBD3 protein.
XX
OS   Xenopus laevis.
XX
FH   Key             Location/Qualifiers
```

```
FT   Binding-site     1..69
FT                    /label= Methyl_CpG_binding_domain
FT   Region           7..9
FT                    /label= Beta1_helix
FT   Region           14..19
FT                    /label= Beta2_helix
FT   Region           20..30
FT                    /label= L1_loop
FT   Region           31..36
FT                    /label= Beta3_helix
FT   Region           40..42
FT                    /label= Beta4_helix
FT   Region           46..52
FT                    /label= Alpha1_helix
FT   Region           53..59
FT                    /label= L2_loop
FT   Region           61..68
FT                    /label= Hairpin_loop
XX
PN   WO200226960-A2.
XX
PD   04-APR-2002.
XX
PF   28-SEP-2001; 2001WO-US42377.
XX
PR   29-SEP-2000; 2000US-236884P.
XX
PA   (SANG-) SANGAMO BIOSCIENCES INC.
XX
PI   Wolffe AP,  Urnov F,  Lai A,  Raschke E;
XX
DR   WPI; 2002-372124/40.
XX
PT   Compartmentalizing a region of interest in cellular chromatin which
PT   facilitates the modulation of the expression of a gene comprises
PT   contacting the gene with a composition comprising a localization domain
PT   and a DNA binding domain -
XX
PS   Example 2; Fig 1B; 85pp; English.
XX
CC   The present invention relates to methods and compositions for regulating
CC   gene expression. In particular the method of compartmentalising a region
CC   of interest in cellular chromatin comprises contacting the region of
CC   interest with a composition that binds to a binding site in cellular
CC   chromatin, where the binding site is in a gene of interest and the
CC   composition comprises a localisation domain (e.g., methyl CpG binding
CC   domain obtained from MECP2, MBD1, MBD2, MBD3, dMBD-like and dMBD-like
CC   delta) and a DNA binding domain (or functional fragment). The method is
CC   useful for compartmentalising a region of interest in cellular chromatin
CC   which facilitates the modulation of the expression of a gene using a
CC   fusion molecule comprising a DNA binding domain and a localisation domain
CC   that binds to the chromatin. The fusion molecules or polypeptides can be
CC   used to prepare pharmaceutical compositions to prevent or treat cancer,
CC   ischaemia, diabetic retinopathy, macular degeneration, HIV infection,
CC   rheumatoid arthritis, psoriasis, sickle cell anaemia, vascular disease,
CC   Alzheimer's disease, muscular dystrophy, neurodegenerative diseases,
CC   cystic fibrosis, stroke, bacterial, viral or fungal infections. Sequences
CC   of the invention are also used in gene therapy. The present sequence is
CC   Xenopus laevis MBD3 LF protein. This sequence is used in the
CC   exemplification of the invention.
XX
SQ   Sequence   282 AA;

  Query Match             78.1%;  Score 1050;  DB 23;  Length 282;
  Best Local Similarity   75.5%;  Pred. No. 1.6e-91;
  Matches  200;  Conservative   33;  Mismatches   26;  Indels    6;  Gaps    2;

Qy       2 DCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNTVDLSSFDFRT 61
           :| ||  ||||||| |:|||||||||||:|||||||||||||||||::|||:|||||
Db       7 ECSAL-QGWKKEEVTRRSGLSAGKSDVYYYSPSGKKFRSKPQLARYLGNSMDLSTFDFRT 65

Qy      62 GKMMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKVKSD 121
           |||: ||: ||:||:| | |||:|||||||| ||:||||||||||||||||||:||||||

Db      66 GKMLMSKINKNRQRMRYDGLNQSKGKPDLNTALPVRQTASIFKQPVTKVTNHPTNKVKSD 125

Qy     122 PQRMNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSAVASALH 181
           ||:  :|||||||||:| ||:| |: |:::|||||||||||||  |||||||:|||||
Db     126 PQKAVDQPRQLFWEKKLSGLNAFDIAEELVKTMELPKGLQGVGPGCTDETLLSAIASALH 185

Qy     182 TSSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEALMAD 241
           ||: |||||:|||||||| |||||||||||||:|||||||||| ||||||||||||||
Db     186 TSTMPITGQLSAAVEKNPGVWLNTSQPLCKAFMVTDEDIRKQEELVQQVRKKLEEALMAD 245

Qy     242 ILSRAADTEE-----MDIEMDSGDE 261
           :|:   :  :     :| ::|  :|
Db     246 MLAHVEEISKDGGAPLDKDIDDEEE 270


RESULT 10
AAE22577
ID   AAE22577 standard; Protein; 303 AA.
XX
AC   AAE22577;
XX
DT   26-JUL-2002  (first entry)
XX
DE   Xenopus laevis MBD3 LF protein.
XX
KW   Gene expression; cellular chromatin; methyl CpG binding domain; cancer;
KW   localisation domain; diabetic retinopathy; ischaemia; HIV infection;
KW   human immuno deficiency virus; macular degeneration; vascular disease;
KW   rheumatoid arthritis; psoriasis; Alzheimer's disease; muscular dystrophy;
KW   sickle cell anaemia; stroke; neurodegenerative disease; cystic fibrosis;
KW   gene therapy; cytostatic; antidiabetic; ophthalmological; vasotropic;
KW   neuroprotective; nootropic; cerebroprotective; antibacterial; antifungal;
KW   antiviral; MBD3 protein.
XX
OS   Xenopus laevis.
XX
FH   Key             Location/Qualifiers
FT   Binding-site    1..90
FT                   /label= Methyl_CpG_binding_domain
FT   Region          7..9
FT                   /label= Beta1_helix
FT   Region          15..21
FT                   /label= Beta2_helix
FT   Region          22..31
FT                   /label= L1_loop
FT   Region          32..37
FT                   /label= Beta3_helix
FT   Region          61..63
FT                   /label= Beta4_helix
FT   Region          67..73
FT                   /label= Alpha1_helix
FT   Region          74..80
FT                   /label= L2_loop
FT   Region          82..89
FT                   /label= Hairpin_loop
XX
PN   WO200226960-A2.
XX
PD   04-APR-2002.
XX
PF   28-SEP-2001; 2001WO-US42377.
XX
PR   29-SEP-2000; 2000US-236884P.
XX
PA   (SANG-) SANGAMO BIOSCIENCES INC.
XX
PI   Wolffe AP,  Urnov F,  Lai A,  Raschke E;
XX
DR   WPI; 2002-372124/40.
XX
PT   Compartmentalizing a region of interest in cellular chromatin which
PT   facilitates the modulation of the expression of a gene comprises
PT   contacting the gene with a composition comprising a localization domain
```

```
PT   and a DNA binding domain -
XX
PS   Example 2; Fig 1B; 85pp; English.
XX
CC   The present invention relates to methods and compositions for regulating
CC   gene expression. In particular the method of compartmentalising a region
CC   of interest in cellular chromatin comprises contacting the region of
CC   interest with a composition that binds to a binding site in cellular
CC   chromatin, where the binding site is in a gene of interest and the
CC   composition comprises a localisation domain (e.g., methyl CpG binding
CC   domain obtained from MECP2, MBD1, MBD2, MBD3, dMBD-like and dMBD-like
CC   delta) and a DNA binding domain (or functional fragment). The method is
CC   useful for compartmentalising a region of interest in cellular chromatin
CC   which facilitates the modulation of the expression of a gene using a
CC   fusion molecule comprising a DNA binding domain and a localisation domain
CC   that binds to the chromatin. The fusion molecules or polypeptides can be
CC   used to prepare pharmaceutical compositions to prevent or treat cancer,
CC   ischaemia, diabetic retinopathy, macular degeneration, HIV infection,
CC   rheumatoid arthritis, psoriasis, sickle cell anaemia, vascular disease,
CC   Alzheimer's disease, muscular dystrophy, neurodegenerative diseases,
CC   cystic fibrosis, stroke, bacterial, viral or fungal infections. Sequences
CC   of the invention are also used in gene therapy. The present sequence is
CC   Xenopus laevis MBD3 LF protein. This sequence is used in the
CC   exemplification of the invention.
XX
SQ   Sequence   303 AA;

  Query Match             77.9%;  Score 1047.5;  DB 23;  Length 303;
  Best Local Similarity   70.5%;  Pred. No. 3e-91;
  Matches  201;  Conservative   33;  Mismatches   26;  Indels   25;  Gaps    2;

Qy       2 DCPALPPGWKKEEVIRKSGLSAGKSDVYYFS--------------------PSGKKFRSK 41
           :| ||| ||||||| |:||||||||||||:|                     |||||||||
Db       7 ECSALPQGWKKEEVTRRSGLSAGKSDVYYYSSSPSRYNRSLRDRVGCLNINPSGKKFRSK 66

Qy      42 PQLARYLGNTVDLSSFDFRTGKMMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTAS 101
           |||||||||::|||:||||||||: ||: ||:||:| | |||:|||||||| ||:|||||
Db      67 PQLARYLGNSMDLSTFDFRTGKMLMSKINKNRQRMRYDGLNQSKGKPDLNTALPVRQTAS 126

Qy     102 IFKQPVTKVTNHPSNKVKSDPQRMNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQ 161
           ||||||||||||:|||||||:  :|||||||||:| ||:| |: |:::|||||||||||
Db     127 IFKQPVTKVTNHPTNKVKSDPQKAVDQPRQLFWEKKLSGLNAFDIAEELVKTMELPKGLQ 186

Qy     162 GVGPGSNDETLLSAVASALHTSSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIR 221
           |||||  |||||||:|||||||: |||||:|||||||| ||||||||||||:|||||||
Db     187 GVGPGCTDETLLSAIASALHTSTMPITGQLSAAVEKNPGVWLNTSQPLCKAFMVTDEDIR 246

Qy     222 KQEERVQQVRKKLEEALMADILSRAADTEE-----MDIEMDSGDE 261
           |||| |||||||||||||||:|:   :  :     :| ::|  :|
Db     247 KQEELVQQVRKKLEEALMADMLAHVEEISKDGGAPLDKDIDDEEE 291


RESULT 11
AAY14198
ID   AAY14198 standard; Protein; 291 AA.
XX
AC   AAY14198;
XX
DT   28-JUL-1999  (first entry)
XX
DE   Human DNA demethylase, dMTase2, protien sequence.
XX
KW   DNA demethylase; dMTase1; dMTase2; human; cancer cell; beta-thalassemia;
KW   DNA methylation pattern; methylation inhibitor; gene therapy; diagnosis;
KW   genetic defect correction; sickle cell anaemia.
XX
OS   Homo sapiens.
XX
PN   WO9924583-A1.
XX
PD   20-MAY-1999.
XX
PF   12-NOV-1998;   98WO-CA01059.
XX
PR   11-MAY-1998;   98CA-2230991.
PR   12-NOV-1997;   97CA-2220805.
XX
PA   (UYMC-) UNIV MCGILL.
XX
PI   Bhattacharya S,  Ramchandani S,  Szyf M;
XX
DR   WPI; 1999-347283/29.
DR   N-PSDB; AAX61219.
XX
PT   Human and murine DNA demethylases useful in the diagnosis of cancer
XX
PS   Disclosure; Fig 9f; 114pp; English.
XX
CC   This sequence is the human DNA demethylase, designated dMTase2, of
CC   the invention. The DNA demethylase is overexpressed in cancer cells.
CC   Expression of the demethylase cDNA is useful to alter DNA methylation
CC   patterns of DNA in vitro in cells or in vivo in humans, animals and
CC   plants. The cDNA is in antisense orientation to inhibit demethylase in
CC   cancer cells for therapeutic purposes. The demethylase is used to alter
CC   the differentiation state and to generate stem cells for therapeutics,
CC   cells for animal cloning and to improve expression of foreign genes. The
CC   cDNA can also be used for recombinant production of large amounts of the
CC   demethylase. The protein can be used to raise antibodies against
CC   demethylase. It can also be used for high throughput screening of
CC   demethylase inhibitors, and for obtaining the x-ray crystal structure.
CC   The demethylase cDNA and protein are also useful for changing the state
CC   of differentiation of a cell to allow gene therapy, stem cell selection
CC   or cell cloning; or for inhibiting methylation in cancer cells using
CC   vector mediated gene therapy. Antagonists or inhibitors of the
CC   demethylase can be used to manufacture medicaments for cancer treatment,
CC   for restoring an aberrant methylation patterns or changing methylation
CC   patterns in patient DNA. Change of the methylation pattern activates a
CC   silent gene permitting the correction of a genetic defect, such as
CC   beta-thalassemia or sickle cell anaemia. The cDNA can be used as a
CC   template to design antisense oligonucleotides and ribozymes. The cDNA can
CC   also be used in two-hybrid systems in yeast to identify proteins
CC   interacting with demethylase. Determining the level of DNA methylase
CC   expression in a cell can be used as a method for diagnosis of cancer,
CC   where overexpression is indicative of cancer cells.
XX
SQ   Sequence   291 AA;

  Query Match             76.0%;  Score 1021;  DB 20;  Length 291;
  Best Local Similarity   72.1%;  Pred. No. 9.6e-89;
  Matches  196;  Conservative   34;  Mismatches   30;  Indels   12;  Gaps    2;

Qy       2 DCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNTVDLSSFDFRT 61
           :||||| ||::||| |:|||||  ||:|:|||||||||||||||| ::|||:|||||
Db       7 ECPALPQGWEREEVPRRSGLSAGHRDVFYYSPSGKKFRSKPQLARYLGGSMDLSTFDFRT 66

Qy      62 GKMMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKVKSD 121
           |||: ||: |::||:| |  || |||||||| ||:|||||||||||||:||||||||||
Db      67 GKMLMSKMNKSRQRVRYDSSNQVKGKPDLNTALPVRQTASIFKQPVTKITNHPSNKVKSD 126

Qy     122 PQRMNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSAVASALH 181
           ||:  :|||||||||:| ||:| |: |:::|||:|||||||||||  |||||||:|||||
Db     127 PQKAVDQPRQLFWEKKLSGLNAFDIAEELVKTMDLPKGLQGVGPGCTDETLLSAIASALH 186

Qy     182 TSSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEALMAD 241
           ||: |||||:|||||||| |||||:|||||||:|||||||||| |||||||:||||||||
Db     187 TSTMPITGQLSAAVEKNPGVWLNTTQPLCKAFMVTDEDIRKQEELVQQVRKRLEEALMAD 246

Qy     242 ILSR----AADTE--------EMDIEMDSGDE 261
           :|:      | | |        | | | |  :|
Db     247 MLAHVEELARDGEAPLDKACAEDDDEEDEEEE 278


RESULT 12
AAY07107
ID   AAY07107 standard; Protein; 200 AA.
XX
```

```
AC   AAY07107;
XX
DT   02-JUL-1999  (first entry)
XX
DE   Colon cancer associated antigen precursor sequence.
XX
KW   Cancer associated antigen; diagnosis; research; treatment; human;
KW   breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW   prostate cancer.
XX
OS   Homo sapiens.
XX
PN   WO9904265-A2.
XX
PD   28-JAN-1999.
XX
PF   15-JUL-1998;   98WO-US14679.
XX
PR   22-JUN-1998;   98US-0102322.
PR   17-JUL-1997;   97US-0896164.
PR   10-OCT-1997;   97US-0061599.
PR   10-OCT-1997;   97US-0061765.
PR   10-OCT-1997;   97US-0948705.
PR   11-OCT-1997;   97GB-0021697.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Chen Y,  Gout I,  Gure A,  O'Hare M,  Obata Y,  Old LJ;
PI   Pfreundschuh M,  Sahin U,  Scanlan MJ,  Stockert E;
PI   Tureci O;
XX
DR   WPI; 1999-132448/11.
XX
PT   New isolated cancer associated nucleic acids and polypeptides -
PT   isolated using sera from cancer patients, used to develop products
PT   for the diagnosis, monitoring or treatment of cancers
XX
PS   Disclosure; Page 677; 787pp; English.
XX
CC   The invention relates to a method for diagnosing a disorder characterised
CC   by expression of a human cancer associated antigen precursor coded for by
CC   a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC   biological sample isolated from a subject with an agent that specifically
CC   binds to the NAM, an expression product or a fragment of an expression
CC   product complexed with an HLA molecule; and (b) determining the
CC   interaction between the agent and the NAM or the expression product as a
CC   determination of the disorder. The products and methods can be used in
CC   the diagnosis, monitoring, research, or treatment of conditions
CC   characterised by the expression of various cancer associated antigens.
CC   The invention provides nucleic acid sequences and encoded polypeptides
CC   which are cancer associated antigen precursors expressed in human breast
CC   cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC   lung cancer.
XX
SQ   Sequence   200 AA;

  Query Match             75.4%;  Score 1013;  DB 20;  Length 200;
  Best Local Similarity   100.0%;  Pred. No. 3.2e-88;
  Matches  200;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      63 KMMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKVKSDP 122
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 KMMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKVKSDP 60

Qy     123 QRMNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSAVASALHT 182
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 QRMNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSAVASALHT 120

Qy     183 SSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEALMADI 242
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 SSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEALMADI 180

Qy     243 LSRAADTEEMDIEMDSGDEA 262
           ||||||||||||||||||||
Db     181 LSRAADTEEMDIEMDSGDEA 200


RESULT 13
AAY14200
ID   AAY14200 standard; Protein; 285 AA.
XX
AC   AAY14200;
XX
DT   28-JUL-1999  (first entry)
XX
DE   Mouse DNA demethylase, dMTase2, protien sequence.
XX
KW   DNA demethylase; dMTase1; dMTase2; mouse; cancer cell; beta-thalassemia;
KW   DNA methylation pattern; methylation inhibitor; gene therapy; diagnosis;
KW   genetic defect correction; sickle cell anaemia.
XX
OS   Mus sp.
XX
PN   WO9924583-A1.
XX
PD   20-MAY-1999.
XX
PF   12-NOV-1998;   98WO-CA01059.
XX
PR   11-MAY-1998;   98CA-2230991.
PR   12-NOV-1997;   97CA-2220805.
XX
PA   (UYMC-) UNIV MCGILL.
XX
PI   Bhattacharya S,  Ramchandani S,  Szyf M;
XX
DR   WPI; 1999-347283/29.
DR   N-PSDB; AAX61221.
XX
PT   Human and murine DNA demethylases useful in the diagnosis of cancer
XX
PS   Disclosure; Fig 9n; 114pp; English.
XX
CC   This sequence is the mouse DNA demethylase, designated dMTase2, of
CC   the invention. The DNA demethylase is overexpressed in cancer cells.
CC   Expression of the demethylase cDNA is useful to alter DNA methylation
CC   patterns of DNA in vitro in cells or in vivo in humans, animals and
CC   plants. The cDNA is in antisense orientation to inhibit demethylase in
CC   cancer cells for therapeutic purposes. The demethylase is used to alter
CC   the differentiation state and to generate stem cells for therapeutics,
CC   cells for animal cloning and to improve expression of foreign genes. The
CC   cDNA can also be used for recombinant production of large amounts of the
CC   demethylase. The protein can be used to raise antibodies against
CC   demethylase. It can also be used for high throughput screening of
CC   demethylase inhibitors, and for obtaining the x-ray crystal structure.
CC   The demethylase cDNA and protein are also useful for changing the state
CC   of differentiation of a cell to allow gene therapy, stem cell selection
CC   or cell cloning; or for inhibiting methylation in cancer cells using
CC   vector mediated gene therapy. Antagonists or inhibitors of the
CC   demethylase can be used to manufacture medicaments for cancer treatment,
CC   for restoring an aberrant methylation patterns or changing methylation
CC   patterns in patient DNA. Change of the methylation pattern activates a
CC   silent gene permitting the correction of a genetic defect, such as
CC   beta-thalassemia or sickle cell anaemia. The cDNA can be used as a
CC   template to design antisense oligonucleotides and ribozymes. The cDNA can
CC   also be used in two-hybrid systems in yeast to identify proteins
CC   interacting with demethylase. Determining the level of DNA methylase
CC   expression in a cell can be used as a method for diagnosis of cancer,
CC   where overexpression is indicative of cancer cells.
XX
SQ   Sequence   285 AA;

  Query Match             75.3%;  Score 1012.5;  DB 20;  Length 285;
  Best Local Similarity   75.3%;  Pred. No. 6e-88;
  Matches  189;  Conservative   35;  Mismatches   24;  Indels    3;  Gaps    1;
```

```
Qy       2 DCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNTVDLSSFDFRT 61
           :|||||  ||::|||  |:|||||   ||:|:|||||||||||||||||| ::|||:|||||
Db       7 ECPALPQGWEREEVPRRSGLSAGHRDVFYYSPSGKKFRSKPQLARYLGGSMDLSTFDFRT 66

Qy      62 GKMMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKVKSD 121
           |||: :|: |::||:| |  || |||||||| ||:|||||||||||||:||||||||||
Db      67 GKMLMNKMNKSRQRVRYDSSNQVKGKPDLNTALPVRQTASIFKQPVTKITNHPSNKVKSD 126

Qy     122 PQRMNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSAVASALH 181
           ||:  :|||||||||:| |||| |: |::::||:|||||||||||  |||||||:|||||
Db     127 PQKAVDQPRQLFWEKKLSGLSAFDIAEELVRTMDLPKGLQGVGPGCTDETLLSAIASALH 186

Qy     182 TSSAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEALMAD 241
           ||: |||||:|||||||| |||||:|||||||:|||:|||||||  |||||||:|||||||
Db     187 TSTLPITGQLSAAVEKNPGVWLNTAQPLCKAFMVTDDDIRKQEELVQQVRKRLEEALMAD 246

Qy     242 ILSRAADTEEEM 252
           :|   |  ||:
Db     247 ML---AHVEEL 254
```

RESULT 14
AAB92997
ID AAB92997 standard; Protein; 223 AA.
XX
AC AAB92997;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:11731.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 11731; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 223 AA;

Query Match 56.0%; Score 753; DB 22; Length 223;
Best Local Similarity 70.5%; Pred. No. 2.1e-63;
Matches 148; Conservative 25; Mismatches 25; Indels 12; Gaps 2;

```
Qy      64 MMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKVKSDPQ 123
           |: ||: |::||:| |  || |||||||| ||:|||||||||||||:|||||||||||
Db       1 MLMSKMNKSRQRVRYDSSNQVKGKPDLNTALPVRQTASIFKQPVTKITNHPSNKVKSDPQ 60

Qy     124 RMNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSAVASALHTS 183
           :  :|||||||||:| ||:| |: |:::|||:|||||||||||  |||||||:|||||||
Db      61 KAVDQPRQLFWEKKLSGLNAFDIAEELVKTMDLPKGLQGVGPGCTDETLLSAIASALHTS 120

Qy     184 SAPITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEALMADIL 243
           : |||||:|||||||| |||||:|||||||:|||||||||||| ||||||:|||||||:|
Db     121 TMPITGQLSAAVEKNPGVWLNTTQPLCKAFMVTDEDIRKQEELVQQVRKRLEEALMADML 180

Qy     244 SR----AADTE--------EMDIEMDSGDE 261
           :      | | |        | | | |  :|
Db     181 AHVEELARDGEAPLDKACAEDDDEEDEEEE 210
```

RESULT 15
AAB93733
ID AAB93733 standard; Protein; 218 AA.
XX
AC AAB93733;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:13377.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 13377; 2537pp + CD ROM; English.

XX
CC The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 218 AA;

Query Match 55.0%; Score 739; DB 22; Length 218;
Best Local Similarity 70.7%; Pred. No. 4.4e-62;
Matches 145; Conservative 24; Mismatches 24; Indels 12; Gaps 2;

```
Qy      69 LQKNKQRLRNDPLNQNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKVKSDPQRMNEQ 128
           : |::||:| |  || |||||||| ||:||||||||||||:||||||||||||:  :|
Db       1 MNKSRQRVRYDSSNQVKGKPDLNTALPVRQTASIFKQPVTKITNHPSNKVKSDPQKAVDQ 60

Qy     129 PRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSAVASALHTSSAPIT 188
           |||||||||:| ||:| |: |:::|||:||||||||||||  |||||||:|||||||: |||
Db      61 PRQLFWEKKLSGLNAFDIAEELVKTMDLPKGLQGVGPGCTDETLLSAIASALHTSTMPIT 120

Qy     189 GQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEALMADILSR--- 245
           ||:|||||||| |||||:|||||||:|||||||||||| ||||||:|||||||:|:
Db     121 GQLSAAVEKNPGVWLNTTQPLCKAFMVTDEDIRKQEELVQQVRKRLEEALMADMLAHVEE 180

Qy     246 -AADTE--------EMDIEMDSGDE 261
            | | |        | | | |  :|
Db     181 LARDGEAPLDKACAEDDDEEDEEEE 205
```

Search completed: March 12, 2003, 05:40:35
Job time : 39.1516 secs

OM protein - protein search, using sw model

Run on: March 12, 2003, 03:11:54 ; Search time 16.4889 Seconds
(without alignments)
1033.836 Million cell updates/sec

Title: US-09-554-414B-2
Perfect score: 2167
Sequence: 1 MRAHPGGGRCCPEQEEGESA.....LSRAADTEEMDIEMDSGDEA 411

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	221.5	10.2	252	1	GRP1_PHAVU	P10495 phaseolus v
2	219.5	10.1	1147	1	MYSB_ACACA	P19706 acanthamoeb
3	214	9.9	641	1	EBN1_EBV	P03211 epstein-bar
4	208.5	9.6	465	1	GRP2_PHAVU	P10496 phaseolus v
5	206.5	9.5	183	1	GRP2_ORYSA	P29834 oryza sativ
6	205.5	9.5	486	1	MEC2_HUMAN	P51608 homo sapien
7	200.5	9.3	165	1	GRP1_ORYSA	P25074 oryza sativ
8	196.5	9.1	481	1	LORI_MOUSE	P18165 mus musculu
9	195.5	9.0	1733	1	VNUA_PRVKA	P33485 pseudorabie
10	192.5	8.9	316	1	LORI_HUMAN	P23490 homo sapien
11	192.5	8.9	384	1	GRP1_PETHY	P09789 petunia hyb
12	192.5	8.9	747	1	SPD1_NEPCL	P19837 nephila cla
13	191.5	8.8	338	1	GRP_ARATH	P27483 arabidopsis
14	191	8.8	386	1	RB87_DROME	P48810 drosophila
15	190	8.8	404	1	CAZ_DROME	Q27294 drosophila
16	190	8.8	1901	1	YZ08_MYCTU	O53553 mycobacteri
17	189.5	8.7	484	1	MEC2_MOUSE	Q9z2d6 mus musculu
18	189.5	8.7	492	1	MEC2_RAT	Q00566 rattus norv
19	188.5	8.7	214	1	GRP2_NICSY	P27484 nicotiana s
20	186	8.6	643	1	K2C1_HUMAN	P04264 homo sapien
21	185	8.5	2038	1	FSH_DROME	P13709 drosophila
22	179.5	8.3	622	1	K1CI_HUMAN	P35527 homo sapien
23	178	8.2	645	1	K22E_HUMAN	P35908 homo sapien
24	177	8.2	593	1	K1CJ_HUMAN	P13645 homo sapien
25	177	8.2	5263	1	FBOH_BOMMO	P05790 bombyx mori
26	176	8.1	629	1	K2C3_HUMAN	P12035 homo sapien
27	175	8.1	1355	1	CA21_RANCA	O42350 rana catesb
28	173.5	8.0	1380	1	DDX9_MOUSE	O70133 mus musculu
29	172	7.9	321	1	FBRL_HUMAN	P22087 homo sapien
30	171	7.9	201	1	GR2B_ARATH	Q38896 arabidopsis
31	170.5	7.9	463	1	YA68_MYCTU	O53416 mycobacteri
32	170	7.8	825	1	5E5_RAT	Q63003 rattus norv
33	169.5	7.8	347	1	MSA2_PLAF2	Q03646 plasmodium
34	169.5	7.8	385	1	RO32_XENLA	P51992 xenopus lae
35	169.5	7.8	491	1	YK98_MYCTU	Q10707 mycobacteri
36	169	7.8	169	1	GR10_BRANA	Q05966 brassica na
37	168.5	7.8	200	1	GRP_HORVU	P17816 hordeum vul
38	167.5	7.7	1356	1	CA21_ONCMY	O93484 oncorhynchu
39	167	7.7	647	1	KNRL_DROME	P13054 drosophila
40	166.5	7.7	526	1	FUS_HUMAN	P35637 homo sapien
41	165	7.6	142	1	GRP1_SORBI	Q99069 sorghum bic
42	165	7.6	401	1	CSP_PLACG	P08674 plasmodium
43	164	7.6	627	1	K2C1_MOUSE	P04104 mus musculu
44	164	7.6	1048	1	AGO1_ARATH	O04379 arabidopsis
45	163.5	7.5	518	1	FUS_MOUSE	P56959 mus musculu

ALIGNMENTS

```
RESULT 1
GRP1_PHAVU
ID   GRP1_PHAVU     STANDARD;      PRT;   252 AA.
AC   P10495;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-JUL-1989 (Rel. 11, Last sequence update)
DT   01-AUG-1992 (Rel. 23, Last annotation update)
DE   Glycine-rich cell wall structural protein 1.0 precursor (GRP 1.0).
OS   Phaseolus vulgaris (Kidney bean) (French bean).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Phaseolus.
OX   NCBI_TaxID=3885;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Tendergreen;
RX   MEDLINE=89091109; PubMed=3208742;
RA   Keller B., Sauer N., Lamb C.J.;
RT   "Glycine-rich cell wall proteins in bean: gene structure and
RT   association of the protein with the vascular system.";
RL   EMBO J. 7:3625-3633(1988).
CC   -!- FUNCTION: RESPONSIBLE FOR PLASTICITY OF THE CELL WALL (POTENTIAL).
CC   -!- SUBCELLULAR LOCATION: Cell wall (Potential).
CC   -!- SIMILARITY: THE N-TERMINAL SIGNAL SEQUENCE OF ABOUT 30 AA SHOWS
CC       ABOUT 60% HOMOLOGY TO THAT OF THE GLYCINE-RICH CELL WALL
CC       STRUCTURAL PROTEIN GRP 1.8.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X13595; CAA31931.1; -.
DR   PIR; S01821; S01821.
KW   Cell wall; Structural protein; Repeat; Signal.
FT   SIGNAL        1     30       POTENTIAL.
FT   CHAIN        31    252       GLYCINE-RICH CELL WALL STRUCTURAL
FT                                PROTEIN 1.0.
FT   DOMAIN       43    252       GLY-RICH.
SQ   SEQUENCE   252 AA;  19958 MW;  831F051691714A00 CRC64;

  Query Match             10.2%;  Score 221.5;  DB 1;  Length 252;
  Best Local Similarity   41.1%;  Pred. No. 1.3e-06;
  Matches   53;  Conservative    7;  Mismatches   64;  Indels    5;  Gaps    2;

Qy      17 GESAAGGSGAGGDSAIEQGGQGSALAPSPVSGVRREGARGGGRGRGRWKQAGRGGGVCGR 76
           |    ||:| ||    : || |          |    || ||| | |     | |||  |
Db      87 GHGGGGGNGGGGGGGADGGGYGGGAGKGGGEGYGGGGANGGGYGGGGGSGGGGGGGAGGA 146

Qy      77 GRGRGRGRGRGRG----RGRGRPPSGGSGLGGDGGGCGGGGSGGG-GAPRREPVPFPS 131
           | | | | | | |     | |   :|| | || ||  ||| :||| || :    :
Db     147 GSGYGGGEGSGAGGGYGGANGGGGGGNGGGGGGGSGGAHGGGAAGGGEGAGQGAGGGYGG 206
```

```
Qy    132 GSAGPGPRG 140
          |:|| | ||
Db    207 GAAGGGGRG 215


RESULT 2
MYSB_ACACA
ID   MYSB_ACACA     STANDARD;      PRT;  1147 AA.
AC   P19706;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Myosin heavy chain IB (Myosin heavy chain IL).
GN   MIB OR MIL.
OS   Acanthamoeba castellanii (Amoeba).
OC   Eukaryota; Acanthamoebidae; Acanthamoeba.
OX   NCBI_TaxID=5755;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90060816; PubMed=2511079;
RA   Jung G., Schmidt C.J., Hammer J.A. III;
RT   "Myosin I heavy-chain genes of Acanthamoeba castellanii: cloning of a
RT   second gene and evidence for the existence of a third isoform.";
RL   Gene 82:269-280(1989).
CC   -!- FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO F-ACTIN & HAS ATPASE
CC       ACTIVITY THAT IS ACTIVATED BY F-ACTIN.
CC   -!- SUBUNIT: MYOSIN I HEAVY CHAIN IS SINGLE-HEADED. DIMER OF A HEAVY
CC       AND A LIGHT CHAIN. INABILITY TO SELF-ASSEMBLE INTO FILAMENTS.
CC   -!- DOMAIN: TH.1 BINDS DIRECTLY TO ANIONIC PHOSPHOLIPID MEMBRANES;
CC       MYOSINS I COULD THEREFORE MOVE ACTIN RELATIVE TO MEMBRANES AND
CC       VICE VERSA. TH.2 AND SH3 BIND TIGHTLY TO F-ACTIN; THIS TOGETHER
CC       WITH THE NUCLEOTIDE-SENSITIVE SITE IN THE HEAD, ALLOWS SINGLE
CC       MOLECULES OF MYOSIN I TO CROSS-LINK ACTIN FILAMENTS.
CC   -!- MISCELLANEOUS: THIS ORGANISM EXPRESSES AT LEAST THREE ISOFORMS OF
CC       MYOSIN I HEAVY-CHAIN, ENCODED BY GENES MIA, MIB, AND MIC.
CC   -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; M30780; AAA27708.1; -.
DR   PIR; JQ0095; MWAXIB.
DR   HSSP; P08799; 1MND.
DR   InterPro; IPR000048; IQ_region.
DR   InterPro; IPR001452; SH3.
DR   InterPro; IPR001609; myosin_head.
DR   Pfam; PF00018; SH3; 1.
DR   Pfam; PF00063; myosin_head; 1.
DR   Pfam; PF00612; IQ; 1.
DR   PRINTS; PR00193; MYOSINHEAVY.
DR   PRINTS; PR00452; SH3DOMAIN.
DR   ProDom; PD000066; SH3; 1.
DR   ProDom; PD000355; myosin_head; 1.
DR   SMART; SM00242; MYSc; 1.
DR   SMART; SM00326; SH3; 1.
DR   PROSITE; PS50002; SH3; 1.
KW   Myosin; ATP-binding; Phosphorylation; Multigene family; SH3 domain.
FT   DOMAIN        1    670       MYOSIN HEAD-LIKE.
FT   DOMAIN      671    907       TAIL HOMOLOGY REGION 1 (TH.1).
FT   DOMAIN      908   1089       GLY/PRO/ALA-RICH (TH.2).
FT   DOMAIN     1090   1147       SH3.
FT   NP_BIND     103    110       ATP (POTENTIAL).
SQ   SEQUENCE   1147 AA;  124958 MW;  B76DE9C076381054 CRC64;

  Query Match             10.1%;  Score 219.5;  DB 1;  Length 1147;
  Best Local Similarity   36.0%;  Pred. No. 9.1e-06;
  Matches   81;  Conservative   12;  Mismatches   85;  Indels   47;  Gaps   12;

Qy      2 RAHPGGGRCCPEQEEGESAAGGSGAGGDSAIEQGGQGSALAPSPVSGVRREGARG---GG 58
          |   ||||       |    || | || |      |  |  | ||: |   |  |   ||
Db    926 RGGGGGGR--GAAGGGRGGFGGGGGGGYSQPVAQAQPVAQVPQPVAAVPSAGRGGPGMGG 983

Qy     59 RGRGRWKQA-GRGG-GVCGRGRGRGRGRGRGRGRGRGRPPSGGSGLGGDGG---GCGG 113
           | ||     |||| |: | | ||| | | |   | ||| |   |:| || ||   | ||
Db    984 PGAGRGGPGMGRGGPGMGGPGAGRG-GPGMGGPGGPGRGGPGGPGAGRGGPGGPGAGRGG 1042

Qy    114 GGSGG-GGAPRREPVPFPSGSAGPGPRGPRATESGKRMDCPALPPGWKKEEVIRKSGLSA 172
          | || ||| |  |    :|  |||  || |   |         ||             |
Db   1043 PGMGGPGGAGRGGP---GAGRGGPGMGGPGAGRGG---------PG------------A 1077

Qy    173 GKSDVYYFSPSGKKFRSKPQLARYLGNTVDLSSFDFRTGKMMPSK 217
          |:      :|:     :|||:         |  :| :||  :  |
Db   1078 GRGAAPAPAPAAP---AKPQVKA-------LYDYDAQTGDELTFK 1112


RESULT 3
EBN1_EBV
ID   EBN1_EBV       STANDARD;      PRT;   641 AA.
AC   P03211;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   EBNA-1 nuclear protein.
GN   BKRF1.
OS   Epstein-barr virus (strain B95-8) (Human herpesvirus 4).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Gammaherpesvirinae; Lymphocryptovirus.
OX   NCBI_TaxID=10377;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=84270667; PubMed=6087149;
RA   Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
RA   Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,
RA   Tuffnell P.S., Barrell B.G.;
RT   "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
RL   Nature 310:207-211(1984).
RN   [2]
RP   SEQUENCE OF 1-26 FROM N.A.
RX   MEDLINE=86259739; PubMed=3460083;
RA   Sample J., Hummel M., Braun D., Birkenbach M., Kieff E.;
RT   "Nucleotide sequences of mRNAs encoding Epstein-Barr virus nuclear
RT   proteins: a probable transcriptional initiation site.";
RL   Proc. Natl. Acad. Sci. U.S.A. 83:5096-5100(1986).
RN   [3]
RP   SUBCELLULAR LOCATION.
RX   MEDLINE=90266473; PubMed=2161150;
RA   Petti L., Sample C., Kieff E.;
RT   "Subnuclear localization and phosphorylation of Epstein-Barr virus
RT   latent infection nuclear proteins.";
RL   Virology 176:563-574(1990).
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF 469-607.
RX   MEDLINE=96006523; PubMed=7553871;
RA   Bochkarev A., Barwell J.A., Pfuetzner R.A., Furey W.F. Jr.,
RA   Edwards A.M., Frappier L.;
RT   "Crystal structure of the DNA-binding domain of the Epstein-Barr
RT   virus origin-binding protein EBNA 1.";
RL   Cell 83:39-46(1995).
CC   -!- FUNCTION: INVOLVED IN LATENT CYCLE. EBNA-1 FUNCTIONS IN THE
CC       MAINTENANCE REPLICATION OF EBV EPISOME. TRANSACTIVATING FACTOR
CC       FOR THE ORIGIN AND ENHANCER FUNCTIONS OF ORIP.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR. FREE IN THE NUCLEOPLASM, SOMEWHAT
CC       ASSOCIATED WITH THE CHROMATIN AND HARDLY, IF AT ALL ASSOCIATED
CC       WITH THE NUCLEAR MATRIX.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
```

```
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; V01555; CAA24816.1; -.
DR   EMBL; M13941; AAA45889.1; -.
DR   PIR; A03773; QQBE31.
DR   PIR; S33021; S33021.
DR   PDB; 1VHI; 23-DEC-96.
DR   TRANSFAC; T00211; -.
DR   InterPro; IPR004186; EBNA1.
DR   Pfam; PF02905; EBNA1; 1.
KW   Nuclear protein; DNA-binding; Transcription regulation; Activator;
KW   3D-structure.
FT   DOMAIN       87    352       GLY/ALA-RICH.
SQ   SEQUENCE   641 AA;  56427 MW;  4D161653E16FC341 CRC64;

  Query Match             9.9%;  Score 214;  DB 1;  Length 641;
  Best Local Similarity   39.7%;  Pred. No. 1e-05;
  Matches   75;  Conservative    1;  Mismatches   57;  Indels   56;  Gaps   11;

Qy       3 AHPGGGRCCPEQEEGESAAGGSGAGGDSAIEQGGQGSALAPSPVSGVRREGARGGGRGRG 62
           |  |||         |   |||:||||      || | | |     |    ||  || |||
Db     278 AGAGGG----AGGAGAGGAGGAGAGGAGGAGAGGAGGAG---GGAGAGGAGAGGGGRG 330

Qy      63 RWKQAGRG-GGVCGRGRG-----RGRG----------RGRGRGRGRGRGRPPSGGSGLGG 106
           |    ||| ||  |||||     ||||          | ||||||||  || |  |
Db     331 RGGSGGRGRGGSGGRGRGGSGGRRGRGRERARGGSRERARGRGRGRGEKRPRSPSS---- 386

Qy     107 DGGGCGGGGSGGGGAPRREPV---PF--PSGSAG-------PGPRG-----PRATESGKR 149
                 |  |  ||| |    ||  | | |          || |     | | | |
Db     387 -------QSSSSGSPPRRPPPGRRPFFHPVGEADYFEYHQEGGPDGEPDVPPGAIEQG-- 437

Qy     150 MDCPALPPG 158
             ||  ||
Db     438 ---PADDPG 443


RESULT 4
GRP2_PHAVU
ID   GRP2_PHAVU     STANDARD;      PRT;   465 AA.
AC   P10496;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-JUL-1989 (Rel. 11, Last sequence update)
DT   01-OCT-1994 (Rel. 30, Last annotation update)
DE   Glycine-rich cell wall structural protein 1.8 precursor (GRP 1.8).
OS   Phaseolus vulgaris (Kidney bean) (French bean).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Phaseolus.
OX   NCBI_TaxID=3885;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Tendergreen;
RX   MEDLINE=89091109; PubMed=3208742;
RA   Keller B., Sauer N., Lamb C.J.;
RT   "Glycine-rich cell wall proteins in bean: gene structure and
RT   association of the protein with the vascular system.";
RL   EMBO J. 7:3625-3633(1988).
CC   -!- FUNCTION: RESPONSIBLE FOR PLASTICITY OF THE CELL WALL (POTENTIAL).
CC   -!- SUBCELLULAR LOCATION: Cell wall (Potential).
CC   -!- DOMAIN: THE REPEATED DOMAINS OF THE PROTEIN FORM A BETA-PLATED
CC       SHEET CONFIGURATION.
CC   -!- SIMILARITY: THE N-TERMINAL SIGNAL SEQUENCE OF ABOUT 30 AA SHOWS
CC       ABOUT 60% HOMOLOGY TO THAT OF THE GLYCINE-RICH CELL WALL
CC       STRUCTURAL PROTEIN GRP 1.0.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X13596; CAA31932.1; -.
DR   PIR; S01820; S01820.
KW   Cell wall; Structural protein; Repeat; Signal.
FT   SIGNAL        1     30       POTENTIAL.
FT   CHAIN        31    465       GLYCINE-RICH CELL WALL STRUCTURAL
FT                                PROTEIN 1.8.
FT   DOMAIN       33    465       GLY-RICH.
FT   DOMAIN      205    359       8 X 22 AA TANDEM REPEATS.
SQ   SEQUENCE   465 AA;  36683 MW;  B5C4A9B983B43607 CRC64;

  Query Match             9.6%;  Score 208.5;  DB 1;  Length 465;
  Best Local Similarity   42.1%;  Pred. No. 1.5e-05;
  Matches   59;  Conservative    5;  Mismatches   53;  Indels   23;  Gaps    6;

Qy       6 GGGRCCPEQEEGESAAGGSGAGGDSAIEQGGQGSALAPSPVSGVRREGARGGGRGRGRWK 65
           |||         :|  | || ||||:  |  || | : |     |    ||:||| | |
Db     120 GGG-------QGGGAGGGYGAGGEHGIGYGGGGGSGAGGG-GGYNAGGAQGGGYGTGGGA 171

Qy      66 -QAGRGGGVCGRGRGRGRGRGRGRGRGRGRPPSGGSGLGGDGGGCGG----GGSGGGG 120
            | |||  | | | |:| | | | | | |    || | || ||| ||    ||  |||
Db     172 GGGGGGGGDHGGGYGGGQGAGGGAGGGYGGGGEHGGGGG-GGQGGGAGGGYGAGGEHGGG 230

Qy     121 APRREPVPFPSGSAGPGPRG 140
           |          |  | |  |
Db     231 A---------GGGQGGGAGG 241


RESULT 5
GRP2_ORYSA
ID   GRP2_ORYSA     STANDARD;      PRT;   183 AA.
AC   P29834;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Glycine-rich cell wall structural protein 2 precursor.
GN   GRP 0.9 OR GRP-1.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. AIJIAO NANTE; TISSUE=Leaf;
RX   MEDLINE=92032791; PubMed=1840687;
RA   Fang R.X., Pang Z., Gao D.M., Mang K.Q., Chua N.H.;
RT   "cDNA sequence of a virus-inducible, glycine-rich protein gene from
RT   rice.";
RL   Plant Mol. Biol. 17:1255-1257(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. AIJIAO NANTE INDICA RICE;
RA   Fang R.X., Wang J.L.;
RL   Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: RESPONSIBLE FOR PLASTICITY OF THE CELL WALL (POTENTIAL).
CC   -!- SUBCELLULAR LOCATION: Cell wall (Potential).
CC   -!- INDUCTION: BY INFECTION WITH RICE YELLOW STUNT VIRUS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X54449; CAA38315.1; -.
DR   EMBL; U40708; AAA85863.1; -.
DR   PIR; S18567; KNRZG2.
KW   Cell wall; Structural protein; Repeat; Signal.
FT   SIGNAL        1     27       POTENTIAL.
FT   CHAIN        28    183       GLYCINE-RICH CELL WALL STRUCTURAL
```

```
FT                                PROTEIN 2.
FT   DOMAIN       28    181       GLY-RICH.
SQ   SEQUENCE   183 AA;  14920 MW;  E9889452101B3893 CRC64;

  Query Match             9.5%;  Score 206.5;  DB 1;  Length 183;
  Best Local Similarity   41.2%;  Pred. No. 7.1e-06;
  Matches   54;  Conservative    5;  Mismatches   53;  Indels   19;  Gaps    3;

Qy       5 PGGGRCCPEQEEGESAAGGSGAGGDSAIEQGGQGSALAPSPVSGVRREGARGGGRGRGRW 64
           ||||             || | ||        | ||        ||    || ||| |||
Db      31 PGGG----------GGGGGEGGGGGYGGSGYGSGSGYGEGGGSG----GAAGGGYGRG-- 74

Qy      65 KQAGRGGGVCGRGRGRGRGRGRGRGRGRGRPPSGGSGLGGDGGGCGGGGSGGGGAPRR 124
             | |||  | | | | | | |:| | | |   :|| | || |||  |||:|| |
Db      75 ---GGGGGGGGEGGGSGSGYGSGQGSGYGAGVGGAGGYGSGGGGGGGQGGGAGGYGQGSG 131

Qy     125 EPVPFPSGSAG 135
             : ||: |
Db     132 YGSGYGSGAGG 142


RESULT 6
MEC2_HUMAN
ID   MEC2_HUMAN     STANDARD;      PRT;   486 AA.
AC   P51608; O15233;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Methyl-CpG-binding protein 2 (MeCP-2 protein) (MeCP2).
GN   MECP2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kudo S., Fukuda M.;
RL   Submitted (SEP-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RA   Thiesen J., Straetling W.H.;
RL   Submitted (APR-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97130625; PubMed=8976388;
RA   Vilain A., Apiou F., Vogt N., Dutrillaux B., Malfoy B.;
RT   "Assignment of the gene for methyl-CpG-binding protein 2 (MECP2) to
RT   human chromosome band Xq28 by in situ hybridization.";
RL   Cytogenet. Cell Genet. 74:293-294(1996).
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Reichwald K., Rosenthal A., Kioschis P., Platzer M.;
RT   "Mapping and sequence analysis of the human MeCP2 locus.";
RL   Submitted (OCT-1997) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99299240; PubMed=10369871;
RA   Coy J.F., Sedlacek Z., Baechner D., Delius H., Poustka A.;
RT   "A complex pattern of evolutionary conservation and alternative
RT   polyadenylation within the long 3'-untranslated region of the
RT   methyl-CpG-binding protein 2 gene (MeCP2) suggests a regulatory role
RT   in gene expression.";
RL   Hum. Mol. Genet. 8:1253-1262(1999).
RN   [6]
RP   SEQUENCE OF 10-486 FROM N.A.
RC   TISSUE=Skeletal muscle;
RX   MEDLINE=96327611; PubMed=8672133;
RA   D'Esposito M., Quaderi N.A., Ciccodicola A., Bruni P., Esposito T.,
RA   D'Urso M., Brown S.D.M.;
RT   "Isolation, physical mapping, and Northern analysis of the X-linked
RT   human gene encoding methyl CpG-binding protein, MECP2.";
RL   Mamm. Genome 7:533-535(1996).
RN   [7]
RP   SEQUENCE OF 10-486 FROM N.A.
RA   Reichwald K., Bauer D., Brenner V., Drescher B., Coy J.,
RA   Kioschis P., Korn B., Nyakatura G., Platzer M., Poustka A.,
RA   Sandoval N., Rosenthal A.;
RL   Submitted (DEC-1996) to the EMBL/GenBank/DDBJ databases.
RN   [8]
RP   VARIANTS RTT TRP-106; CYS-133; SER-155 AND MET-158.
RX   MEDLINE=99438392; PubMed=10508514;
RA   Amir R.E., Van den Veyver I.B., Wan M., Tran C.Q., Francke U.,
RA   Zoghbi H.Y.;
RT   "Rett syndrome is caused by mutations in X-linked MECP2, encoding
RT   methyl-CpG-binding protein 2.";
RL   Nat. Genet. 23:185-188(1999).
RN   [9]
RP   VARIANT RTT VAL-140.
RX   MEDLINE=20465115; PubMed=11007980;
RA   Orrico A., Lam C., Galli L., Dotti M.T., Hayek G., Tong S.F.,
RA   Poon P.M., Zappella M., Federico A., Sorrentino V.;
RT   "MECP2 mutation in male patients with non-specific X-linked mental
RT   retardation.";
RL   FEBS Lett. 481:285-288(2000).
RN   [10]
RP   VARIANTS RTT W-106; F-124; C-13; C-134; R-152; M-158 AND C-306.
RX   MEDLINE=20439334; PubMed=10991688;
RA   Obata K., Matsuishi T., Yamashita Y., Fukuda T., Kuwajima K.,
RA   Horiuchi I., Nagamitsu S., Iwanaga R., Kimura A., Omori I., Endo S.,
RA   Mori K., Kondo I.;
RT   "Mutation analysis of the methyl-CpG binding protein 2 gene (MECP2) in
RT   patients with Rett syndrome.";
RL   J. Med. Genet. 37:608-610(2000).
RN   [11]
RP   VARIANTS RTT R-101; W-106; M-158 AND C-306, AND VARIANT K-397.
RX   MEDLINE=20439335; PubMed=10991689;
RA   Hampson K., Woods C.G., Latif F., Webb T.;
RT   "Mutations in the MECP2 gene in a cohort of girls with Rett
RT   syndrome.";
RL   J. Med. Genet. 37:610-612(2000).
RN   [12]
RP   VARIANT RTT VAL-140.
RX   MEDLINE=21664240; PubMed=11805248;
RA   Dotti M.T., Orrico A., De Stefano N., Battisti C., Sicurelli F.,
RA   Severi S., Lam C.W., Galli L., Sorrentino V., Federico A.;
RT   "A Rett syndrome MECP2 mutation that causes mental retardation in
RT   men.";
RL   Neurology 58:226-230(2002).
CC   -!- FUNCTION: CHROMOSOMAL PROTEIN THAT BINDS TO METHYLATED DNA. IT CAN
CC       BIND SPECIFICALLY TO A SINGLE METHYL-CPG PAIR. IT IS NOT
CC       INFLUENCED BY SEQUENCES FLANKING THE METHYL-CPGS. MEDIATES
CC       TRANSCRIPTIONAL REPRESSION THROUGH INTERACTION WITH HISTONE
CC       DEACETYLASE AND THE COREPRESSOR SIN3A.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR. COLOCALIZED WITH METHYL-CPG IN THE
CC       GENOME.
CC   -!- TISSUE SPECIFICITY: PRESENT IN ALL ADULT SOMATIC TISSUES TESTED.
CC   -!- DISEASE: DEFECTS IN MECP2 ARE THE CAUSE OF RETT SYNDROME (RTT), AN
CC       X-LINKED DOMINANT DISEASE. RTT IS A PROGRESSIVE NEUROLOGIC
CC       DEVELOPMENTAL DISORDER AND ONE OF THE MOST COMMON CAUSES OF MENTAL
CC       RETARDATION IN FEMALES. PATIENTS APPEAR TO DEVELOP NORMALLY UNTIL
CC       6 TO 18 MONTHS OF AGE, THEN GRADUALLY LOOSE SPEECH AND PURPOSEFUL
CC       HAND MOVEMENTS AND DEVELOP MICROCEPHALY, SEIZURES, AUTISM, ATAXIA,
CC       INTERMITTENT HYPERVENTILATION, AND STEREOTYPIC HAND MOVEMENTS.
CC       AFTER INITIAL REGRESSION, THE CONDITION STABILIZES AND PATIENTS
CC       USUALLY SURVIVE INTO ADULTHOOD.
CC   -!- SIMILARITY: CONTAINS 1 A.T HOOK DNA-BINDING REPEAT.
CC   -!- SIMILARITY: CONTAINS 1 METHYL-BINDING DOMAIN (MBD).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
```

```
CC   ----------------------------------------------------------------------
DR   EMBL; L37298; AAC32737.1; -.
DR   EMBL; Y12643; CAA73190.1; -.
DR   EMBL; X99686; CAA68001.1; -.
DR   EMBL; AF030876; AAC08757.1; -.
DR   EMBL; AF031078; AAC08758.1; -.
DR   EMBL; AJ132917; CAB46446.1; -.
DR   EMBL; X89430; CAA61599.1; -.
DR   EMBL; X94628; CAA64331.1; -.
DR   TRANSFAC; T04936; -.
DR   Genew; HGNC:6990; MECP2.
DR   MIM; 300005; -.
DR   MIM; 312750; -.
DR   InterPro; IPR000637; AT_hook.
DR   InterPro; IPR001739; Methyl-CpG_bind.
DR   Pfam; PF01429; MBD; 1.
DR   SMART; SM00384; AT_hook; 1.
DR   SMART; SM00391; MBD; 1.
KW   Transcription regulation; Repressor; DNA-binding; Nuclear protein;
KW   Disease mutation; Polymorphism.
FT   DOMAIN       96    149       MBD.
FT   DOMAIN      277    283       POLY-ALA.
FT   DOMAIN      366    372       POLY-HIS.
FT   DOMAIN      384    393       POLY-PRO.
FT   VARIANT     101    101       P -> R (IN RTT).
FT                                /FTId=VAR_010276.
FT   VARIANT     106    106       R -> W (IN RTT).
FT                                /FTId=VAR_010272.
FT   VARIANT     124    124       L -> F (IN RTT).
FT                                /FTId=VAR_010277.
FT   VARIANT     133    133       R -> C (IN RTT).
FT                                /FTId=VAR_010273.
FT   VARIANT     134    134       S -> C (IN RTT).
FT                                /FTId=VAR_010278.
FT   VARIANT     140    140       A -> V (IN RTT; AFFECTS WOMEN AND MEN).
FT                                /FTId=VAR_010279.
FT   VARIANT     152    152       P -> R (IN RTT).
FT                                /FTId=VAR_010280.
FT   VARIANT     155    155       F -> S (IN RTT).
FT                                /FTId=VAR_010274.
FT   VARIANT     158    158       T -> M (IN RTT).
FT                                /FTId=VAR_010275.
FT   VARIANT     201    201       A -> V.
FT                                /FTId=VAR_010281.
FT   VARIANT     306    306       R -> C (IN RTT).
FT                                /FTId=VAR_010282.
FT   VARIANT     397    397       E -> K.
FT                                /FTId=VAR_010283.
FT   CONFLICT     72     75       PAVP -> RLC (IN REF. 6).
FT   CONFLICT    290    290       E -> G (IN REF. 3).
SQ   SEQUENCE   486 AA;  52440 MW;  EB6A33233AEDA566 CRC64;

  Query Match             9.5%;  Score 205.5;  DB 1;  Length 486;
  Best Local Similarity  28.5%;  Pred. No. 2.4e-05;
  Matches   85;  Conservative   38;  Mismatches  104;  Indels   71;  Gaps   15;

Qy     117 GGGGAPRREPVPFPSGSAGPGPRGPRATESGKRMDCPALPPGWKKEEVIRKSGLSAGKSD 176
           | | ||       |  || |  |     : |   | | || || ::   |||| |||| |
Db      67 GSGSAP-----AVPEASASPKQRRSIIRDRGPMYDDPTLPEGWTRKLKQRKSGRSAGKYD 121

Qy     177 VYYFSPSGKKFRSKPQLARY---LGNT-VDLSSFDFR-TGKMMPSKLQKNKQRLRNDP-- 229
           ||  :| || |||| :|  |   :|:| :| : |||  ||:  ||: ::   :    |
Db     122 VYLINPQGKAFRSKVELIAYFEKVGDTSLDPNDFDFTVTGRGSPSRREQKPPKKPKSPKA 181

Qy     230 --LNQNKGKP-DLNTTLPIRQTASIFKQPVTKVTNHPSNKVKSDPQRMNEQPRQLFWEKR 286
             : :|:|    || |   |:   :  | :|        ||  : : : | |     :
Db     182 PGTGRGRGRPKGSGTTRPKAATSEGVQ--VKRVLE------KSPGKLLVKMPFQTSPGGK 233

Qy     287 LQGLSASDVTE-QIIK----------TMELPKGLQGVGPGSNDETLLSAVASALHTSSA 334
           :|  |:  |:  :||             :||  :|  ||       | ||:|
Db     234 AEGGGATTSTQVMVIKRPGRKRKAEADPQAIPK-KRGRKPG-------SVVAAA------ 279

Qy     335 PITGQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRK-KLEEALMADI 391
```

```
               ||  |  |             | :  ||  :| |  ::| |  | :  ::
Db     280 -------AAEAKKKA-------------VKESSIRSVQETVLPIKKRKTRETVSIEV 316


RESULT 7
GRP1_ORYSA
ID   GRP1_ORYSA     STANDARD;      PRT;   165 AA.
AC   P25074;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   01-APR-1993 (Rel. 25, Last annotation update)
DE   Glycine-rich cell wall structural protein 1 precursor.
GN   GRP-1.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Indica-IR36;
RX   MEDLINE=91370862; PubMed=1716496;
RA   Lei M., Wu R.;
RT   "A novel glycine-rich cell wall protein gene in rice.";
RL   Plant Mol. Biol. 16:187-198(1991).
CC   -!- FUNCTION: RESPONSIBLE FOR PLASTICITY OF THE CELL WALL (POTENTIAL).
CC   -!- SUBCELLULAR LOCATION: Cell wall (Potential).
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X53596; CAA37665.1; -.
DR   PIR; S13385; KNRZG1.
KW   Cell wall; Structural protein; Repeat; Signal.
FT   SIGNAL        1     23       POTENTIAL.
FT   CHAIN        24    165       GLYCINE-RICH CELL WALL STRUCTURAL
FT                                PROTEIN 1.
FT   DOMAIN       31    159       GLY-RICH.
FT   REPEAT       56     62       R2 (TYR-RICH).
FT   REPEAT       93     99       R2 (TYR-RICH).
FT   REPEAT      132    138       R2 (TYR-RICH).
SQ   SEQUENCE   165 AA;  13536 MW;  E36CE31C3650AC9A CRC64;

  Query Match             9.3%;  Score 200.5;  DB 1;  Length 165;
  Best Local Similarity  40.8%;  Pred. No. 1.4e-05;
  Matches   51;  Conservative    6;  Mismatches   59;  Indels    9;  Gaps    3;

Qy      17 GESAAGGSGAGGDSAIEQGGQGSALAPSPVSGVRREGAR-GGGRGRGRWKQAGRGGGVCG 75
           | |  || | ||       || ||        |  : | : |||:| |     | |||  |
Db      32 GGSGGGGGGGGGGGGG---GGNGSGSGSGYGYGYGKAGGQSGGGQGSG-----GGGGGGGG 83

Qy      76 RGRGRGRGRGRGRGRGRGRGRPPSGGSGLGGDGGGCGGGGSGGGGAPRREPVPFPSGSAG 135
           | | | | | | | | |:| |     ||| || ||| ||||  | |:       :  |  |
Db      84 GGNGSGSGSGYGYGYGQGNGGAQGQGSGGGGGGGGGGGGGGSGQGSGSGYGYGYGKGGGG 143

Qy     136 PGPRG 140
           |  |
Db     144 GGGGG 148


RESULT 8
LORI_MOUSE
ID   LORI_MOUSE     STANDARD;      PRT;   481 AA.
AC   P18165;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Loricrin.
```

```
GN   LOR.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90275605; PubMed=2190691;
RA   Mehrel T., Hohl D., Rothnagel J.A., Longley M.A., Bundman D.,
RA   Cheng C., Lichti U., Bisher M.E., Steven A.C., Steinart P.M.,
RA   Yuspa S.H., Roop D.R.;
RT   "Identification of a major keratinocyte cell envelope protein,
RT   loricrin.";
RL   Cell 61:1103-1112(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/c;
RX   MEDLINE=95256248; PubMed=7738016;
RA   DiSepio D., Jones A., Longley M.A., Bundman D., Rothnagel J.A.,
RA   Roop D.R.;
RT   "The proximal promoter of the mouse loricrin gene contains a
RT   functional AP-1 element and directs keratinocyte-specific but not
RT   differentiation-specific expression.";
RL   J. Biol. Chem. 270:10792-10799(1995).
CC   -!- FUNCTION: MAJOR KERATINOCYTE CELL ENVELOPE PROTEIN.
CC   -!- SUBUNIT: MONOMERS ARE CROSSLINKED BY DISULFIDE AND N-(GAMMA-
CC       GLUTAMYL) LYSINE ISODIPEPTIDE BONDS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M34398; AAA39444.1; -.
DR   EMBL; U09189; AAA82152.1; -.
DR   PIR; A35628; A35628.
DR   HSSP; P02876; 9WGA.
DR   MGD; MGI:96816; Lor.
KW   Keratinization.
SQ   SEQUENCE   481 AA;  37830 MW;  97349A786FF239FE CRC64;

  Query Match             9.1%;  Score 196.5;  DB 1;  Length 481;
  Best Local Similarity   36.7%;  Pred. No. 7.8e-05;
  Matches   61;  Conservative    4;  Mismatches   60;  Indels   41;  Gaps    6;

Qy       6 GGGRCCPEQEEGESAAGGSGAGGDSAIEQGGQGS--------------------ALAPS 44
           ||  |       | |: |||| || |    || ||                      : ||
Db     305 GGSSCGGGSSGGGSSCGGSGGGGYSG---GGGGSCGGGSSGGGGGYYSSQQTSQTSCAPQ 361

Qy      45 PVSGVRREGARGGGRGRGRWKQAGRGGVCGRGRGRGRGRG---RGRGRGRGPPSGG 101
               |    |: |||   |     | ||| |    | |   | |    | | | |   |||
Db     362 QSYG---GGSSGGGGSCGGGSSGGGGGGGCYSSGGGGSSGGCGGGYSGGGGGCGGGSSGG 418

Qy     102 SG-------LGGDGGGCGGGGSGGGGAPRREPVPFPSGSAGPGPRG 140
           ||        || ||||||| |||||          ||:| |  |
Db     419 SGGGCGGGSSGGSGGGCGGGYSGGGGGGS----SCGGGSSGGGSGG 460


RESULT 9
VNUA_PRVKA
ID   VNUA_PRVKA     STANDARD;      PRT;  1733 AA.
AC   P33485;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   01-FEB-1994 (Rel. 28, Last annotation update)
DE   Probable nuclear antigen.
OS   Pseudorabies virus (strain Kaplan) (PRV).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Varicellovirus.
OX   NCBI_TaxID=33703;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91021039; PubMed=2171211;
RA   Vlcek C., Kozmik Z., Paces V., Schirm S., Schwyzer M.;
RT   "Pseudorabies virus immediate-early gene overlaps with an oppositely
RT   oriented open reading frame: characterization of their promoter and
RT   enhancer regions.";
RL   Virology 179:365-377(1990).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M34651; AAA47471.1; -.
DR   PIR; B45344; B45344.
FT   DOMAIN      112    117       POLY-THR.
FT   DOMAIN      179   1733       GLY-RICH.
FT   DOMAIN      192    196       POLY-SER.
FT   DOMAIN      271    298       POLY-PRO.
FT   DOMAIN      304    308       POLY-ARG.
FT   DOMAIN      883    889       POLY-GLY.
FT   DOMAIN     1398   1405       POLY-GLY.
SQ   SEQUENCE   1733 AA;  172166 MW;  0C8CD8BE475BB5E2 CRC64;

  Query Match             9.0%;  Score 195.5;  DB 1;  Length 1733;
  Best Local Similarity   25.4%;  Pred. No. 0.00036;
  Matches   99;  Conservative   30;  Mismatches  149;  Indels  111;  Gaps   19;

Qy      12 PEQEEGESAAGGSGAGGD---SAIEQG--GQGSALAPSP--------VSGVRREGARG-- 56
           ||:        || | ||    | :  |   :|:|||| |           |  ||  |
Db     873 PERAPEPVVLGGGGGGGQQRGSGVRSGPESEGAALAPGPPVLFVVAVAVAVPAEGRAGEP 932

Qy      57 ---------GGRGR--------GRWKQAGRGG-GVCGRGRGRGRGRGRGRGRGRGRPP 98
                    | ||         ||| :|| || || |     | | | | |
Db     933 LVLLAVPGAAGPGRAALLLAPLGRWVRAGGGGAGVAGGAGEAGLGAGAGLG--------- 983

Qy      99 SGGSGLGGDGGGCGGGGSGGGGAPRR-------------EPVPFPSGSAGPGPRG---- 140
             |:|||  | |  | |  |||| ||                |      |  ||||||
Db     984 -AGAGLGAGGAGGPGAGEAGGGARRRRRRRWDDEAGLLGPERGQAGRGLRGPGPRGGLGE 1042

Qy     141 PRATESGKRMDCPALPPGWKKEEVIRKSGLSAGKSDVYYFS--------PSGKKFRSKPQ 192
           |     |:  :   : ||          || ||   |:  :          | : |  |
Db    1043 PGRGHVGRGEEGRGVGPG----------GL-AGAGPVHAVAHQRRHGAGDEGDRVRGLPP 1091

Qy     193 LARYLGNTVDLSSFDFRTGKMMPSKLQKNKQRLR-NDPLN----QNKGKPDLNTTLPIRQ 247
           | |                | :  : |: :  |    |        :|:|:      :
Db    1092 LGR------------AGPGDRVAEREQRGRHLLEAGGPEGGRGAGGRGQPERAGQQALED 1139

Qy     248 TASIFKQPVTKVTNHPSNKVKSDPQRMNEQPRQLFWEKRLQGLSASDVTEQIIKTMELPK 307
            |:     | ::  | :  ::      :       |  : || |
Db    1140 AAAGQDAGVRQLAGHAAG-LRGGEGGADAGAEGL--DGRLPGAGVRGAAR---------V 1187

Qy     308 GLQGVGPGS--NDETLLSAVASALHTSSA 334
           |  ||||     || || |: :| |   |
Db    1188 GHVGVGPAEALQDEGLLGAIVAAAHGHGA 1216


RESULT 10
LORI_HUMAN
ID   LORI_HUMAN     STANDARD;      PRT;   316 AA.
AC   P23490;
DT   01-NOV-1991 (Rel. 20, Created)
DT   01-NOV-1991 (Rel. 20, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Loricrin.
GN   LOR OR LRN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=Foreskin;
RX   MEDLINE=91177926; PubMed=2007607;
RA   Hohl D., Mehrel T., Lichti U., Turner M.L., Roop D.R., Steinert P.M.;
RT   "Characterization of human loricrin. Structure and function of a new
RT   class of epidermal cell envelope proteins.";
RL   J. Biol. Chem. 266:6626-6636(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92388173; PubMed=1355480;
RA   Yoneda K., Hohl D., McBride O.W., Wang M., Cehrs K.U., Idler W.W.,
RA   Steinert P.M.;
RT   "The human loricrin gene.";
RL   J. Biol. Chem. 267:18060-18066(1992).
CC   -!- FUNCTION: MAJOR KERATINOCYTE CELL ENVELOPE PROTEIN.
CC   -!- SUBUNIT: MONOMERS ARE CROSSLINKED BY DISULFIDE AND N-(GAMMA-
CC       GLUTAMYL) LYSINE ISODIPEPTIDE BONDS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M61120; AAA36180.1; -.
DR   EMBL; M94077; AAA36181.1; -.
DR   PIR; A38743; A38743.
DR   Genew; HGNC:6663; LOR.
DR   MIM; 152445; -.
KW   Keratinization.
SQ   SEQUENCE   316 AA;  26033 MW;  100FEC52FD9ACE8E CRC64;

  Query Match             8.9%;  Score 192.5;  DB 1;  Length 316;
  Best Local Similarity   32.3%;  Pred. No. 8.5e-05;
  Matches   65;  Conservative    7;  Mismatches   68;  Indels   61;  Gaps    9;

Qy       6 GGGRCCPEQEEGESAAGGSG----------AGGDSAIEQGG---QGSALAPSPVSGVRRE 52
           |||  |      | |: ||||           ||      ||    | |:      ||
Db     108 GGGSGCFSSGGGGSSGGGSGCFSSGGGGSSGGGSGCFSSGGGGFSGQAVQCQSYGGVSSG 167

Qy      53 GARGGGRGRGRWKQAGRGGGVC---GRGRGRGRGRGRGRGRGRG-------------- 95
           |: ||  | | :   | || ||   | | | | | | |   | | |
Db     168 GSSGG--GSGCFSSGGGGGSVCGYSGGGSGGGSGCGGGSSGGSGSGYVSSQQVTQTSCAP 225

Qy      96 RPPSGG--SGLGGDGG------------GCGGGGSG--------GGGAPRREPVPFPSG 132
           :|  ||  || || ||             ||||| ||        |||:          |
Db     226 QPSYGGGSSGGGGSGGSGCFSSGGGGGSSGCGGGSSGIGSGCIISGGGS------VCGGG 279

Qy     133 SAGPGPRGPRATESGKRMDCP 153
           |:| |  |     ||     |
Db     280 SSGGGGGGSSVGGSGSGKGVP 300


RESULT 11
GRP1_PETHY
ID   GRP1_PETHY     STANDARD;      PRT;   384 AA.
AC   P09789;
DT   01-MAR-1989 (Rel. 10, Created)
DT   01-MAR-1989 (Rel. 10, Last sequence update)
DT   01-AUG-1992 (Rel. 23, Last annotation update)
DE   Glycine-rich cell wall structural protein 1 precursor.
GN   GRP-1.
OS   Petunia hybrida (Petunia).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Asteridae; euasterids I; Solanales; Solanaceae; Petunia.
OX   NCBI_TaxID=4102;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Condit C.M., Meagher R.B.;
RT   "A gene encoding a novel glycine-rich structural protein of petunia.";
RL   Nature 323:178-181(1986).
CC   -!- FUNCTION: RESPONSIBLE FOR PLASTICITY OF THE CELL WALL (POTENTIAL).
CC   -!- SUBCELLULAR LOCATION: Cell wall (Potential).
CC   -!- MISCELLANEOUS: THIS PROTEIN CONTAINS 67% GLYCINE RESIDUES.
CC   -!- MISCELLANEOUS: 90% OF THE MATURE PROTEIN RESIDUES ARE CAPABLE OF
CC       FORMING A BETA-PLEATED SHEET COMPOSED OF 8 ANTI-PARALLEL STRANDS.
CC   -!- MISCELLANEOUS: THE GLYCINE-RICH REGION IS COMPRISED OF TWO RELATED
CC       FAMILIES OF REPEATS, F1 AND F2, EACH REPEAT CONTAINING ABOUT 40
CC       AA.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X04335; CAA27866.1; -.
DR   PIR; A26099; A26099.
KW   Cell wall; Structural protein; Repeat; Signal.
FT   SIGNAL        1     27
FT   CHAIN        28    384       GLYCINE-RICH CELL WALL STRUCTURAL
FT                                 PROTEIN 1.
FT   DOMAIN       41    384       GLY-RICH.
SQ   SEQUENCE   384 AA;  28777 MW;  C8541C549417D18C CRC64;

  Query Match             8.9%;  Score 192.5;  DB 1;  Length 384;
  Best Local Similarity   40.4%;  Pred. No. 0.00011;
  Matches   55;  Conservative    3;  Mismatches   55;  Indels   23;  Gaps    5;

Qy       6 GGGRCCPEQEEGESAAGGSGAGGDSAIEQGGQGSALAPSPVSGVRREGARGGGRGRGRWK 65
           |||         |    || | || :    || |          |:   |  ||| | |
Db      91 GGG--------GAGGGGGLGGGGGAG---GGFGGGAGGGAGGGLGGGGGLGGGGGGG--- 136

Qy      66 QAGRGGGV-CGRGRGRGRGRGRGRGRGRGRPPSGGSGLGGDGGGCGGGGSGGGGAPRR 124
             || ||||  | | | | | | | | | | | |    || | || |||  ||||| ||
Db     137 -AGGGGGVGGGAGSGGGFGAGGGVGGGAGAGGGVGGGGGFGGGGGGGVGGGSGHGGG--- 192

Qy     125 EPVPFPSGSAGPGPRG 140
             | :|    |  |
Db     193 ----FGAGGGVGGGAG 204


RESULT 12
SPD1_NEPCL
ID   SPD1_NEPCL     STANDARD;      PRT;   747 AA.
AC   P19837;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Spidroin 1 (Dragline silk fibroin 1) (Fragment).
OS   Nephila clavipes (Orb spider).
OC   Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC   Araneomorphae; Entelegynae; Araneoidea; Tetragnathidae; Nephila.
OX   NCBI_TaxID=6915;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=90384959; PubMed=2402494;
RA   Xu M., Lewis R.V.;
RT   "Structure of a protein superfiber: spider dragline silk.";
RL   Proc. Natl. Acad. Sci. U.S.A. 87:7120-7124(1990).
RN   [2]
RP   SEQUENCE OF 653-747 FROM N.A.
RX   MEDLINE=94165058; PubMed=8120021;
RA   Beckwitt R., Arcidiacono S.;
RT   "Sequence conservation in the C-terminal region of spider silk
RT   proteins (Spidroin) from Nephila clavipes (Tetragnathidae) and
RT   Araneus bicentenarius (Araneidae).";
RL   J. Biol. Chem. 269:6661-6663(1994).
CC   -!- FUNCTION: Spiders major ampullate silk possesses unique
```

```
CC       characteristics of strength and elasticity. Fibroin consists of
CC       pseudocristalline regions of antiparallel beta-sheet interspersed
CC       with elastic amorphous segments.
CC   -!- SUBUNIT: MAJOR SUBUNIT, WITH SPIDROIN 2, OF THE DRAGLINE SILK.
CC   -!- SUBCELLULAR LOCATION: Extracellular.
CC   -!- DOMAIN: Highly repetitive protein characterized by regions of
CC       polyalanine and glycine-rich repeating units.
CC   -!- SIMILARITY: BELONGS TO THE SILK FIBROIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M37137; AAA29380.1; -.
DR   EMBL; U03848; AAB60212.1; -.
DR   PIR; A36068; A36068.
KW   Silk; Repeat.
FT   NON_TER       1      1
FT   DOMAIN        1    655       25 X APPROXIMATE TANDEM REPEATS.
FT   REPEAT        1     25       1.
FT   REPEAT       26     38       2.
FT   REPEAT       39     66       3.
FT   REPEAT       67     96       4.
FT   REPEAT       97    130       5.
FT   REPEAT      131    158       6.
FT   REPEAT      159    191       7.
FT   REPEAT      192    204       8.
FT   REPEAT      205    235       9.
FT   REPEAT      236    262       10.
FT   REPEAT      263    292       11.
FT   REPEAT      293    305       12.
FT   REPEAT      306    333       13.
FT   REPEAT      334    360       14.
FT   REPEAT      361    394       15.
FT   REPEAT      395    424       16.
FT   REPEAT      425    458       17.
FT   REPEAT      459    485       18.
FT   REPEAT      486    512       19.
FT   REPEAT      513    525       20.
FT   REPEAT      526    555       21.
FT   REPEAT      556    582       22.
FT   REPEAT      583    612       23.
FT   REPEAT      613    642       24.
FT   REPEAT      643    655       25.
FT   CONFLICT    662    662       V -> L (IN REF. 1).
FT   CONFLICT    672    672       S -> T (IN REF. 1).
FT   CONFLICT    695    747       NPGLSGCDVLIQALLEVVSALIQILGSSSIGQVNYGSAGQA
FT                                TQIVGQSVYQAL -> ILVFLDVMSSFKLFSRLFLLLSRS
FT                                (IN REF. 1).
SQ   SEQUENCE   747 AA;  60528 MW;  850E44B0D649E012 CRC64;

  Query Match             8.9%;  Score 192.5;  DB 1;  Length 747;
  Best Local Similarity   30.0%;  Pred. No. 0.00022;
  Matches   62;  Conservative   22;  Mismatches   56;  Indels   67;  Gaps    9;

Qy       6 GGGRCCPEQEEGESAAGGSGAG-----GDSAIEQGGQGSALAPSPVSGVRREGARGGGRG 60
           | ||     :   :||||:| |     |     :||||:  | :   |  :|| || | |
Db     458 GAGR---GGQGAAAAAGGAGQGGYGGLGSQGAGRGGQGAGAAAAAAVGAGQEGIRGQGAG 514

Qy      61 RGRW-----KQAGRG------------------GGVCGRGRGRGRG------------- 83
           :| :     : :|||                  ||: |:| |:| |
Db     515 QGGYGGLGSQGSGRGGLGGQGAGAAAAAAGGAGQGGLGGQGAGQGAGAAAAAAGGVRQGG 574

Qy      84 ------RGRGR-GRGRGRGRPPSGGSGLGGDGG----GCGGGGSGGGGAPRREPVPFPSG 132
                 :| || |:| |     :||:| || ||    | | || || ||         :|
Db     575 YGGLGSQGAGRGGQGAGAAAAAAGGAGQGGYGGLGGQGVGRGGLGGQGAGAA-----AAG 629

Qy     133 SAGPGPRG------PRATESGKRMDCP 153
             || |  |       |: :  |:  |
```

```
Db     630 GAGQGGYGGVGSGASAASAAASRLSSP 656


RESULT 13
GRP_ARATH
ID   GRP_ARATH      STANDARD;      PRT;   338 AA.
AC   P27483;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   01-AUG-1992 (Rel. 23, Last annotation update)
DE   Glycine-rich cell wall structural protein precursor.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=92003708; PubMed=1912511;
RA   Quigley F., Villiot M.L., Mache R.;
RT   "Nucleotide sequence and expression of a novel glycine-rich protein
RT   gene from Arabidopsis thaliana.";
RL   Plant Mol. Biol. 17:949-952(1991).
CC   -!- FUNCTION: RESPONSIBLE FOR PLASTICITY OF THE CELL WALL (POTENTIAL).
CC   -!- SUBCELLULAR LOCATION: Cell wall (Potential).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X58338; CAA41249.1; -.
DR   PIR; S17732; KNMU.
KW   Cell wall; Structural protein; Repeat; Signal.
FT   SIGNAL        1     20       POTENTIAL.
FT   CHAIN        21    338       GLYCINE-RICH CELL WALL STRUCTURAL
FT                                PROTEIN.
FT   DOMAIN       21    338       GLY-RICH.
SQ   SEQUENCE   338 AA;  23891 MW;  046A6E8C1A4E89EB CRC64;

  Query Match             8.8%;  Score 191.5;  DB 1;  Length 338;
  Best Local Similarity   40.9%;  Pred. No. 0.0001;
  Matches   61;  Conservative    2;  Mismatches   53;  Indels   33;  Gaps    8;

Qy       6 GGGRCCPEQEEGESAAGGSGAGGDSAIE-------QGGQGSALAPSPVSGVRREGARGGG 58
           |||         |  | || ||||   :          || |  |       |   | |||
Db     193 GGG-------AGGGAGGGGGAGGGGGLGGGHGGGFGGGAGGGLGGGAGGGT--GGGFGGG 243

Qy      59 RGRGRWKQAGRG-GGVCGRGRGRGRGRGRGRGRGRGRPPSGGSGLGGDGGGCG---GG 114
             | |    || | ||  | | | | | | | | | | | |    ||:| || ||| |   ||
Db     244 AGGGAGGGAGGGFGGGAGGGAGGGFGGGAGGGAGGGAGGGFGGGAG-GGHGGGVGGGFGG 302

Qy     115 GSG---GGGAPRREPVPFPSGSAGPGPRG 140
           |||   ||||          | || |  |
Db     303 GSGGGFGGGA---------GGGAGGGAGG 322


RESULT 14
RB87_DROME
ID   RB87_DROME     STANDARD;      PRT;   386 AA.
AC   P48810;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Heterogeneous nuclear ribonucleoprotein 87F (HRP36.1 protein) (P11
DE   protein).
GN   HRB87F OR HRP36.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
```

```
OC   Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC   Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Oregon-R, and Canton-S; TISSUE=Ovary;
RX   MEDLINE=91187645; PubMed=1849257;
RA   Haynes S.R., Johnson D., Raychaudhuri G., Beyer A.L.;
RT   "The Drosophila Hrb87F gene encodes a new member of the A and B hnRNP
RT   protein group.";
RL   Nucleic Acids Res. 19:25-31(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Canton-S;
RX   MEDLINE=92112968; PubMed=1730754;
RA   Matunis E.L., Matunis M.J., Dreyfuss G.;
RT   "Characterization of the major hnRNP proteins from Drosophila
RT   melanogaster.";
RL   J. Cell Biol. 116:257-269(1992).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Canton-S; TISSUE=Embryo;
RX   MEDLINE=92020124; PubMed=1717937;
RA   Hovemann B.T., Dessen E., Mechler H., Mack E.;
RT   "Drosophila snRNP associated protein P11 which specifically binds to
RT   heat shock puff 93D reveals strong homology with hnRNP core protein
RT   A1.";
RL   Nucleic Acids Res. 19:4909-4914(1991).
CC   -!- FUNCTION: THIS PROTEIN IS A COMPONENT OF RIBONUCLEOSOMES. COULD BE
CC       NEEDED TO ORGANIZE A CONCENTRATION GRADIENT OF A DORSALIZING
CC       MORPHOGEN (DM) ORIGINATING IN THE GERMINAL VESICLE.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR AND/OR CYTOPLASMIC.
CC   -!- SIMILARITY: CONTAINS 2 RNA RECOGNITION MOTIFS (RRM).
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; X54803; CAA38574.1; -.
DR   EMBL; X62636; CAA44502.1; -.
DR   EMBL; X59691; CAA42212.1; -.
DR   HSSP; P09651; 1UP1.
DR   FlyBase; FBgn0004237; Hrb87F.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00076; rrm; 2.
DR   SMART; SM00360; RRM; 2.
DR   PROSITE; PS50102; RRM; 2.
DR   PROSITE; PS00030; RRM_RNP_1; 2.
KW   RNA-binding; Nuclear protein; Ribonucleoprotein; Repeat;
KW   Alternative splicing.
FT   DOMAIN       24    101       RNA-BINDING (RRM) 1.
FT   DOMAIN      115    192       RNA-BINDING (RRM) 2.
FT   VARSPLIC    315    374       MISSING (IN ISOFORM HRP36.1).
FT   CONFLICT    271    271       S -> T (IN REF. 3).
SQ   SEQUENCE   386 AA;  39557 MW;  2036C04D01E3AFD7 CRC64;

  Query Match             8.8%;  Score 191;  DB 1;  Length 386;
  Best Local Similarity  36.2%;  Pred. No. 0.00013;
  Matches   55;  Conservative    8;  Mismatches   67;  Indels   22;  Gaps    6;

Qy       6 GGGRCCPEQ----EEGESAAGGSGAGGDSAIEQGGQGSALAPSPVSGVRREGAR-GGGR- 59
           ||||  |      :|:   || | || :   | | |:         |    |   |||:
Db     202 GGGRGGPRAGGRGGQGDRGQGGGGWGGQN--RQNGGGNWGGAGGGGGFGNSGGNFGGGQG 259

Qy      60 -GRGRWKQAGRGGGVCGRGRGRGRGRGRGRGRGRGRPPSGGSGLGGDGGGCGGGGSGG 118
            | | | | |  ||     :| | |   | | | | |   | ||  |  ||| |||| |
Db     260 GGSGGWNQQGGSGGGPWNNQGGGNGGWNGGGGGGGYGGGNSNGSWGGNGGGGGGGGGFGN 319

Qy     119 ------GGAPRR-------EPVPFPSGSAGPG 137
                 || |:|        | |:  |  | |
Db     320 EYQQSYGGGPQRNSNFGNNRPAPYSQGGGGGG 351


RESULT 15
CAZ_DROME
ID   CAZ_DROME      STANDARD;      PRT;   404 AA.
AC   Q27294; Q24445; Q9VXI4;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   RNA-binding protein cabeza (Sarcoma-associated RNA-binding fly
DE   homolog) (P19).
GN   CAZ OR SARFH OR CG3606.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC   Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC   Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC   STRAIN=Canton-S;
RX   MEDLINE=95349623; PubMed=7623847;
RA   Immanuel D., Zinszner H., Ron D.;
RT   "Association of SARFH (sarcoma-associated RNA-binding fly homolog)
RT   with regions of chromatin transcribed by RNA polymerase II.";
RL   Mol. Cell. Biol. 15:4562-4571(1995).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   STRAIN=Canton-S;
RX   MEDLINE=95223793; PubMed=7708500;
RA   Stolow D.T., Haynes S.R.;
RT   "Cabeza, a Drosophila gene encoding a novel RNA binding protein,
RT   shares homology with EWS and TLS, two genes involved in human sarcoma
RT   formation.";
RL   Nucleic Acids Res. 23:835-843(1995).
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
```

```
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [4]
RP   SEQUENCE OF 39-404 FROM N.A.
RC   STRAIN=Oregon-R;
RA   Haynes S.R.;
RL   Submitted (APR-1988) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE OF 212-261 FROM N.A.
RC   STRAIN=Oregon-R;
RX   MEDLINE=87175568; PubMed=3031652;
RA   Haynes S.R., Rebbert M.L., Mozer B.A., Forquignon F., Dawid I.B.;
RT   "Pen repeat sequences are GGN clusters and encode a glycine-rich
RT   domain in a Drosophila cDNA homologous to the rat helix destabilizing
RT   protein.";
RL   Proc. Natl. Acad. Sci. U.S.A. 84:1819-1823(1987).
CC   -!- FUNCTION: MAY PARTICIPATE IN A FUNCTION COMMON TO THE EXPRESSION
CC       OF MOST GENES TRANSCRIBED BY RNA POLYMERASE II.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; 1 AND 2 (SHOWN HERE);
CC       MAY BE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: UBIQUITOUS. ENRICHED IN THE BRAIN AND CENTRAL
CC       NERVOUS SYSTEM DURING EMBRYOGENESIS. ENRICHED IN THE ADULT HEAD.
CC       EMBRYOS CONTAIN BOTH TYPE 1 AND TYPE 2 ISOFORMS, WHEREAS LATER IN
CC       DEVELOPMENT (HEADS AND TORSOS) ONLY THE TYPE 2 ISOFORM IS
CC       DETECTED.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED IN THE DEVELOPING EMBRYO FROM THE
CC       EARLIEST STAGES OF CELLULARIZATION AND IS SUBSEQUENTLY FOUND IN
CC       MANY CELL TYPES.
CC   -!- MISCELLANEOUS: 'CABEZA' MEANS 'HEAD' IN SPANISH.
CC   -!- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
CC   -!- SIMILARITY: CONTAINS 1 RANBP2-TYPE ZINC FINGER.
CC   -!- SIMILARITY: BELONGS TO THE TET FAMILY OF RNP PROTEINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U13178; AAA86955.1; -.
DR   EMBL; L37083; AAC41563.1; -.
DR   EMBL; AE003501; AAF48578.1; -.
DR   EMBL; M15765; AAA70425.1; -.
DR   FlyBase; FBgn0011571; caz.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   InterPro; IPR001876; Znf_RanGDP.
DR   Pfam; PF00076; rrm; 1.
DR   Pfam; PF00641; zf-RanBP; 1.
DR   SMART; SM00360; RRM; 1.
DR   SMART; SM00547; ZnF_RBZ; 1.
DR   PROSITE; PS50102; RRM; 1.
DR   PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.
DR   PROSITE; PS01358; ZF_RANBP2_1; 1.
DR   PROSITE; PS50199; ZF_RANBP2_2; 1.
KW   Nuclear protein; Zinc-finger; Metal-binding; RNA-binding;
KW   Alternative splicing.
FT   DOMAIN       42    111       GLY-RICH.
FT   DOMAIN      119    205       RNA-BINDING (RRM).
FT   DOMAIN      212    275       GLY-RICH.
FT   ZN_FING     280    309       RANBP2-TYPE.
FT   DOMAIN      312    391       GLY-RICH.
FT   VARSPLIC      4     47       MISSING (IN ISOFORM 1).
FT   CONFLICT     39     41       PNY -> LFI (IN REF. 4).
FT   CONFLICT     92     92       P -> H (IN REF. 3).
FT   CONFLICT    108    108       G -> GG (IN REF. 3).
FT   CONFLICT    253    258       MISSING (IN REF. 3).
FT   CONFLICT    283    283       D -> E (IN REF. 4 AND 5).
FT   CONFLICT    389    398       DGGPMRNDGG -> MVDQEKRWS (IN REF. 4).
SQ   SEQUENCE   404 AA;  39141 MW;  7062A0446BEA5984 CRC64;

  Query Match             8.8%;  Score 190;  DB 1;  Length 404;
  Best Local Similarity   30.3%;  Pred. No. 0.00016;
  Matches   59;  Conservative    4;  Mismatches   68;  Indels   64;  Gaps    5;

Qy       6 GGGRCCPEQEEGESAAGGSGAGGDSAIEQGGQGSALAPSPVSGVRREGARGGGRGRGR-- 63
           ||||        |    || | ||    :::|| |              | |  ||| | |
Db     219 GGGRGGFGGRRGGGGGGGGGGGGGGRFDRGGGGGGNGGGGGGRYDRGGGGGGGGGGNVQ 278

Qy      64 -----WK----------------------------QAGRGGGVCGRGRGRGRGRGRGRG 89
                 ||                            | |||  | | | |  ||  ||
Db     279 PRDGDWKCNSCNNTNFAWRNECNRCKTPKGDDEGSSGGGGGGGYGGGGGGGGYDRGNDRG 338

Qy      90 RG------RGRGRPPSGGSGLGGDGGGC---------GGGGSGGGGAPRREPVPFPSGSA 134
           |       | ||    || | || |||          | || ||||  ||:
Db     339 SGGGGYHNRDRGGNSQGGGGGGGGGGGYSRFNDNNGGGRGGRGGGGGNRRD--------- 389

Qy     135 GPGPRGPRATESGKR 149
             ||   : | |
Db     390 ----GGPMRNDGGMR 400


Search completed: March 12, 2003, 08:44:08
Job time : 28.4889 secs
```

OM protein - protein search, using sw model

Run on: March 12, 2003, 03:22:03 ; Search time 29.9242 Seconds
(without alignments)
1320.377 Million cell updates/sec

Title: US-09-554-414B-2
Perfect score: 2167
Sequence: 1 MRAHPGGGRCCPEQEEGESA.....LSRAADTEEMDIEMDSGDEA 411

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	225	10.4	592	2	E82759	endo-1,4-beta-gluc
2	221.5	10.2	252	1	S01821	glycine-rich prote
3	220	10.2	221	2	T04592	glycine-rich cell
4	219.5	10.1	1147	1	MWAXIB	myosin heavy chain
5	217.5	10.0	385	2	T20410	hypothetical prote
6	214	9.9	271	2	S34666	glycine-rich prote
7	214	9.9	641	1	QQBE31	nuclear antigen EB
8	211.5	9.8	291	1	S31415	glycine-rich prote
9	208.5	9.6	396	2	T49109	glycine-rich prote
10	208.5	9.6	465	1	S01820	glycine-rich cell
11	207	9.6	255	2	B84777	hypothetical prote
12	206.5	9.5	183	1	KNRZG2	glycine-rich cell
13	204.5	9.4	1585	2	T31611	hypothetical prote
14	202.5	9.3	302	2	C84470	hypothetical prote
15	200.5	9.3	165	1	KNRZG1	glycine-rich cell
16	200.5	9.3	173	2	JQ1064	glycine-rich prote
17	200	9.2	210	2	JQ1060	glycine-rich prote
18	198.5	9.2	408	2	S57483	glycin-rich protei
19	198	9.1	667	2	A70893	hypothetical glyci
20	197.5	9.1	207	2	T07381	glycine-rich prote
21	196.5	9.1	481	2	A35628	loricrin - mouse
22	195.5	9.0	1733	1	B45344	probable nuclear a
23	195	9.0	588	2	F70971	hypothetical glyci
24	194	9.0	1660	2	A70869	hypothetical glyci
25	193.5	8.9	183	2	PN0109	keratin-like prote
26	193.5	8.9	476	2	S57963	methyl CpG binding
27	193.5	8.9	718	2	A36068	major ampullate fi
28	192.5	8.9	316	1	A38743	loricrin - human
29	192.5	8.9	384	1	A26099	glycine-rich cell
30	191.5	8.8	338	1	KNMU	glycine-rich cell
31	191	8.8	386	1	S22315	snRNP-associated p
32	190.5	8.8	171	2	H84709	probable glycine-r
33	190.5	8.8	1215	2	T32734	myosin-IA - Acanth
34	190	8.8	404	2	S54729	RNA-binding protei
35	190	8.8	1901	2	F70806	hypothetical glyci
36	189.5	8.7	492	2	A41907	methyl-CpG-binding
37	189	8.7	594	2	G70545	hypothetical glyci
38	189	8.7	1958	2	B40505	hypothetical prote
39	188.5	8.7	214	1	KNNT2S	glycine-rich prote
40	188	8.7	220	2	T14441	glycine-rich prote
41	187.5	8.7	1079	2	B70807	hypothetical glyci
42	186.5	8.6	270	2	A60830	keratin, 70k type
43	186.5	8.6	1329	2	E70917	hypothetical glyci
44	186	8.6	798	2	T33022	hypothetical prote
45	186	8.6	1300	2	T03166	probable immediate

ALIGNMENTS

RESULT 1
E82759
endo-1,4-beta-glucanase XF0818 [imported] - Xylella fastidiosa (strain 9a5c)
C;Species: Xylella fastidiosa
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C;Accession: E82759
R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A;Reference number: A82515; MUID:20365717; PMID:10910347
A;Note: for a complete list of authors see reference number A59328 below
A;Accession: E82759
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-592 <SIM>
A;Cross-references: GB:AE003921; GB:AE003849; NID:g9105710; PIDN:AAF83628.1; GSPDB:GN
A;Experimental source: strain 9a5c
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; La
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.
, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri,
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawa
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv
M.; Tsuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.
A;Reference number: A59328
A;Contents: annotation
C;Genetics:
A;Gene: XF0818

Query Match 10.4%; Score 225; DB 2; Length 592;
Best Local Similarity 29.4%; Pred. No. 6.8e-07;
Matches 92; Conservative 20; Mismatches 101; Indels 100; Gaps 14;

```
Qy      6 GGGRCCPEQEEGESAAGGSGAGGDSAIEQGGQGSALAPSPVSGVRREGARGGGRGRGRWK 65
          |||        |  : ||||:|| |    || ||       ||       ||| | |
Db    378 GGG-----SGGGSGSGGGSGSGGGSG-SGGGSGSGGG----SGSGGGSGSGGGSGSGGGS 427

Qy     66 QAGRGGGVCGRGRGRGRGRGRGRGRGRGRPPSGGSGLGGDGGGCGGGGSGGGGAPRRE 125
          :| |||  | | | | | | | | | |    |||| || ||| |||||||||
Db    428 GSGGGGG-SGGGSGSGGGSGSGGGSGSGGGSGSGGGSGSGG-GGGSGGGGSGGGG----- 480

Qy    126 PVPFPSGSAGPGPRGPRATESGKRMDCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGK 185
              || |    | :   |                    ||   | ||
Db    481 ------GSGGGSGSGGGSGSGG------------------GSGTGGGGSD--------- 506

Qy    186 KFRSKPQLARYLGNTVDLSSFDFRTGKMMPSKLQKNKQRLRNDPLNQNKGKPDLNTTLPI 245
              |  ::   ||:  |  |: ||           :|::
```

```
Db    507 -----PTKSKISVNTIVDS--DWHTGYC---------ERIQ------------------- 531

Qy    246 RQTASIFKQPVTKVTNHPSNKVKSDPQRMNEQPR-QLFWEKRLQGLSASDVTEQIIKTME 304
                  || ::| |:    :   | | |      |  :   |:|| :     ||::
Db    532 ----------VTNISNSPTTWTVTISIRGNIQTLWSARWSVKGNALTASGMDWN--KTLD 579

Qy    305 LPKGLQGVGPGSN 317
           |||    |  ||
Db    580 -PKGTTEFGFCSN 591


RESULT 2
S01821
glycine-rich protein 1.0 precursor - kidney bean
C;Species: Phaseolus vulgaris (kidney bean)
C;Date: 30-Sep-1989 #sequence_revision 19-May-1994 #text_change 16-Jul-1999
C;Accession: S01821
R;Keller, B.; Sauer, N.; Lamb, C.J.
EMBO J. 7, 3625-3633, 1988
A;Title: Glycine-rich cell wall proteins in bean: gene structure and association of the
A;Reference number: S01820; MUID:89091109; PMID:3208742
A;Accession: S01821
A;Molecule type: DNA
A;Residues: 1-252 <KEL>
A;Cross-references: EMBL:X13595; NID:g21000; PIDN:CAA31931.1; PID:g21001
C;Superfamily: Phaseolus glycine-rich protein 1.0
C;Keywords: cell wall; structural protein
F;1-30/Domain: signal sequence #status predicted <SIG>
F;31-252/Product: glycine-rich protein 1.0 #status predicted <MAT>

  Query Match             10.2%;  Score 221.5;  DB 1;  Length 252;
  Best Local Similarity   41.1%;  Pred. No. 4e-07;
  Matches   53;  Conservative    7;  Mismatches   64;  Indels    5;  Gaps    2;

Qy     17 GESAAGGSGAGGDSAIEQGGQGSALAPSPVSGVRREGARGGGRGRGRWKQAGRGGGVCGR 76
          |    ||:| ||    : || |         |    || ||| | |     | |||  |
Db     87 GHGGGGGNGGGGGGGADGGGYGGGAGKGGGEGYGGGGANGGGYGGGGGSGGGGGGGAGGA 146

Qy     77 GRGRGRGRGRGRG----RGRGRPPSGGSGLGGDGGGCGGGGSGGG-GAPRREPVPFPS 131
          | | | | | | | |     | |   :|| | || ||  ||| :||| || :     :
Db    147 GSGYGGGEGSGAGGGYGGANGGGGGGNGGGGGGGSGGAHGGGAAGGGEGAGQGAGGGYGG 206

Qy    132 GSAGPGPRG 140
          |:|| | ||
Db    207 GAAGGGGRG 215


RESULT 3
T04592
glycine-rich cell wall structural protein homolog F23E13.120 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 14-May-1999
C;Accession: T04592
R;Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Jesse, T.
submitted to the Protein Sequence Database, March 1998
A;Reference number: Z15378
A;Accession: T04592
A;Molecule type: DNA
A;Residues: 1-221 <BEV>
A;Cross-references: EMBL:AL022141
A;Experimental source: cultivar Columbia; BAC clone F23E13
C;Genetics:
A;Map position: 4
A;Note: F23E13.120

  Query Match             10.2%;  Score 220;  DB 2;  Length 221;
  Best Local Similarity   40.5%;  Pred. No. 4.2e-07;
  Matches   62;  Conservative    5;  Mismatches   50;  Indels   36;  Gaps    5;

Qy     15 EEGES-AAGGSGAGGDSAIEQGGQGSALAPSPVSGVRREGARGGGRGRG---------RW 64
          |: || | || | ||      |||||               ||| | |         |
Db     82 EKCESRAGGGGGGGGGGGGGGGQGS-------------GGGGGEGNGGNGKDNSHKRN 127

Qy     65 KQAGRGGGVCGRGRGRGRGRGRGRGRGRGRPPSGGSGLGGDGGGCGGGGSGGGGAPRR 124
          | :| |||  | | | | | | |||||| | |    || | || ||| |||| |||   :
Db    128 KSSGGGGG-GGGGGGGGSGNGSGRGRGGGGGGGGGGGGGGGGGGGGGGGGGGSDGKG 186

Qy    125 EPVPFPSGSAGPGPRG----------PRATES 146
             |  |  | || |            ::|||
Db    187 GGWGFGFGWGGDGPGGNFGRCWYPGCTKKSTES 219


RESULT 4
MWAXIB
myosin heavy chain IB - Acanthamoeba castellanii
N;Contains: myosin ATPase (EC 3.6.4.1)
C;Species: Acanthamoeba castellanii
C;Date: 30-Sep-1990 #sequence_revision 30-Sep-1990 #text_change 19-Apr-2002
C;Accession: JQ0095; B34448
R;Jung, G.; Schmidt, C.J.; Hammer III, J.A.
Gene 82, 269-280, 1989
A;Title: Myosin I heavy-chain genes of Acanthamoeba castellanii: cloning of a second
A;Reference number: JQ0095; MUID:90060816; PMID:2511079
A;Accession: JQ0095
A;Molecule type: DNA
A;Residues: 1-1147 <JUN>
A;Cross-references: GB:M30780
A;Note: this organism expresses at least three isoforms of myosin I heavy-chain, enco
of gene MIB, whereas the protein previously identified as MIB is the product of gene
R;Brzeska, H.; Lynch, T.J.; Martin, B.; Korn, E.D.
J. Biol. Chem. 264, 19340-19348, 1989
A;Title: The localization and sequence of the phosphorylation sites of Acanthamoeba m
A;Reference number: A34448; MUID:90037074; PMID:2530230
A;Accession: B34448
A;Molecule type: protein
A;Residues: 538-550,'X',552-559 <BRZ>
C;Comment: In this protein, the coiled-coil rod-like region found in many myosin heav
protein is globular and does not self-assemble into filaments.
C;Genetics:
A;Gene: MIB
A;Introns: 1/3; 39/3; 102/2; 135/3; 183/3; 212/1; 291/3; 379/3; 492/3; 617/2; 649/3;
C;Superfamily: protozoan myosin heavy chain IB; myosin motor domain homology; SH3 hom
C;Keywords: actin binding; ATP; hydrolase; nucleotide binding; P-loop; phosphoprotein
F;12-664/Domain: myosin motor domain homology <MMOT>
F;103-110/Region: nucleotide-binding motif A (P-loop)
F;552-573/Region: actin binding #status predicted
F;698-1147/Domain: carboxyl-terminal <CTD>
F;698-889/Region: basic
F;910-1094/Region: alanine/glycine/proline-rich
F;1097-1144/Domain: SH3 homology <SH3>
F;109/Binding site: ATP (Lys) #status predicted
F;315/Binding site: phosphate (Ser) (covalent) #status predicted

  Query Match             10.1%;  Score 219.5;  DB 1;  Length 1147;
  Best Local Similarity   36.0%;  Pred. No. 3.4e-06;
  Matches   81;  Conservative   12;  Mismatches   85;  Indels   47;  Gaps   12;

Qy      2 RAHPGGGRCCPEQEEGESAAGGSGAGGDSAIEQGGQGSALAPSPVSGVRREGARG---GG 58
          |   ||||        |    || | || |       |  |  | ||: |   |  |   ||
Db    926 RGGGGGGR--GAAGGGRGGFGGGGGGGYSQPVAQAQPVAQVPQPVAAVPSAGRGGPGMGG 983

Qy     59 RGRGRWKQA-GRGG-GVCGRGRGRGRGRGRGRGRGRGRPPSGGSGLGGDGG---GCGG 113
           | ||     |||| |: | | ||| | | |   | ||| |   |:| || ||   | ||
Db    984 PGAGRGGPGMGRGGPGMGGPGAGRG-GPGMGGPGGPGRGGPGGPGAGRGGPGGPGAGRGG 1042

Qy    114 GGSGG-GGAPRREPVPFPSGSAGPGPRGPRATESGKRMDCPALPPGWKKEEVIRKSGLSA 172
           | || ||| |  |    :|  |||  || |   |          ||             |
Db   1043 PGMGGPGGAGRGGP---GAGRGGPGMGGPGAGRGG---------PG-------------A 1077

Qy    173 GKSDVYYFSPSGKKFRSKPQLARYLGNTVDLSSFDFRTGKMMPSK 217
          |:      :|:     :|||:          |  :| :||  :  |
Db   1078 GRGAAPAPAPAAP---AKPQVKA-------LYDYDAQTGDELTFK 1112


RESULT 5
T20410
```

```
hypothetical protein E02A10.2 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T20410
R;Thomas, K.
submitted to the EMBL Data Library, October 1996
A;Reference number: Z19271
A;Accession: T20410
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-385 <WIL>
A;Cross-references: EMBL:Z81053; PIDN:CAB02877.1; GSPDB:GN00023; CESP:E02A10.2
A;Experimental source: clone E02A10
C;Genetics:
A;Gene: CESP:E02A10.2
A;Map position: 5
A;Introns: 32/1; 72/1; 85/1; 122/1; 133/1; 220/3

  Query Match             10.0%;  Score 217.5;  DB 2;  Length 385;
  Best Local Similarity   23.1%;  Pred. No. 1.2e-06;
  Matches  100;  Conservative   35;  Mismatches  137;  Indels  161;  Gaps   15;

Qy       6 GGGRCCPEQEEGESAAGGSGAGGDSAIEQGGQGSALAPSPVSGVRREGARGGGRGRGRWK 65
           |||  |           || | ||      || |   || |        | ||| | | |
Db      81 GGGGC-------GGGGGGCGGGGGGCGGGGGCGGGCAPPP-----PPPACGGGCGGG--- 125

Qy      66 QAGRGGGVCGRGRGRGRGRGRGRGRGRGRPPSGGSGLGGDGGGCGGGGSGGGGAPRRE 125
            |  || || | | | | | | | | | |    || | || |||||||| || | |
Db     126 -GGGCGGGCGGGGGGGCGGGGGGGCGGGGGGCGGGGGGCGGGGGGCGGGGGGCG----- 179

Qy     126 PVPFPSGSAGPGPRGPRATESGKRMDCPALPPGWKKEEVIRKSG-------------LSA 172
                | | |  |           |     | ||.| |   |              :
Db     180 ------GGGGGGCGG----------GC-----GRKKREAINVVGHDDLNKCNNEELRIIL 218

Qy     173 GKSDVYYFSPSGKKFRSK-----------PQLARYLGNTVDLSSFDFRTGKMMPSKLQKN 221
           ::     : | |  : |            : | :   | |  |             |||
Db     219 NENTKETVAESIKTLKEKVAGQDYIVVCNEKPAPFTAETDDFCS-----------LQKE 266

Qy     222 KQRLRNDPLNQ----NKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKVKSDPQRMNEQ 277
                 :|      | : |     | ::     |: | |          | : :: :|:
Db     267 NVHCTILRINHKEVAEKNEEDKKEEEPKKEEEK--KEEVEK---------KEEDEKKDEE 315

Qy     278 PRQLFWEKRLQGLSASDVTEQIIKTMELPKGLQGVGPGSNDETLLSAVASALHTSSAPIT 337
           |      |: :      :  |:: |  |  |
Db     316 P------KKEEEKKEEEQKEEVEKKEEEEK---------------------------- 339

Qy     338 GQVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEALMADILSRAAD 397
                                    ||: :|:||: ::  || :|     :  ::
Db     340 --------------------------KDEEPKKEEEKKEEEEKKEDE-----VEEKSEK 367

Qy     398 TEEMDIEMDSGDE 410
           || ::|    :|
Db     368 VEEKELEPKKDEE 380


RESULT 6
S34666
glycine-rich protein - common tobacco
C;Species: Nicotiana tabacum (common tobacco)
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 20-Aug-1999
C;Accession: S34666
R;Brady, K.; Darvill, A.G.; Albersheim, P.
submitted to the EMBL Data Library, July 1993
A;Description: Activation of a tobacco glycine-rich protein gene by a fungal glucan prep
A;Reference number: S34666
A;Accession: S34666
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-271 <BRA>
A;Cross-references: EMBL:X74106; NID:g395146; PIDN:CAA52208.1; PID:g395147
C;Superfamily: Phaseolus glycine-rich protein 1.0

  Query Match             9.9%;  Score 214;  DB 2;  Length 271;
  Best Local Similarity   43.4%;  Pred. No. 1.3e-06;
  Matches   59;  Conservative    4;  Mismatches   43;  Indels   30;  Gaps    7;

Qy       6 GGGRCCPEQEEGESAAG-GSGAGGDSAIEQGGQGSALAPSPVSGVRREGARGGGRGRGRW 64
           ||||       ||   | | | ||   :  || |          |    ||:||| | |
Db     159 GGGR----GGGGEGGGGVGGGGGGHGGVRGGGGGGQ------GGGGHGGAQGGGGGYG-- 206

Qy      65 KQAGRGGGVCGRGRGRGRGRGRGRGRGRGRPPSGGSGLGGDGGGCGGGGSGGGGAPRR 124
             ||||| | |:| | | |  :| ||| |    || | || ||| |||| ||
Db     207 ---GRGGG--GGGQG-GGGAGGAQGGGRGGGGVGGGGGGAGGRGGGQGGGGGAGG----- 255

Qy     125 EPVPFPSGSAGPGPRG 140
                 || || | ||
Db     256 ------SGGAGGGGRG 265


RESULT 7
QQBE31
nuclear antigen EBNA1 - human herpesvirus 4
C;Species: human herpesvirus 4, Epstein-Barr virus
C;Date: 25-Feb-1985 #sequence_revision 25-Feb-1985 #text_change 22-Oct-1999
C;Accession: C43043; S42440; A03773; S33021
R;Bankier, A.T.; Deininger, P.L.; Farrell, P.J.; Barrell, B.G.
Mol. Biol. Med. 1, 21-45, 1983
A;Title: Sequence analysis of the 17,166 bp EcoRI fragment C of B95-8 Epstein-Barr vi
A;Reference number: A93065; MUID:85035713; PMID:6092825
A;Accession: C43043
A;Molecule type: DNA
A;Residues: 1-641 <BAN>
A;Cross-references: EMBL:V01555; NID:g59074; PIDN:CAA24816.1; PID:g1334880
A;Experimental source: strain B95-8
R;Baer, R.; Bankier, A.T.; Biggin, M.D.; Deininger, P.L.; Farrell, P.J.; Gibson, T.J.
Nature 310, 207-211, 1984
A;Title: DNA sequence and expression of the B95-8 Epstein-Barr virus genome.
A;Reference number: A03794; MUID:84270667; PMID:6087149
A;Contents: annotation; protein coding region
R;Sample, J.; Hummel, M.; Braun, D.; Birkenbach, M.; Kieff, E.
Proc. Natl. Acad. Sci. U.S.A. 83, 5096-5100, 1986
A;Title: Nucleotide sequences of mRNAs encoding Epstein-Barr virus nuclear proteins:
A;Reference number: S42440; MUID:86259739; PMID:3460083
A;Accession: S42440
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-66 <SAM>
A;Cross-references: EMBL:M13941; NID:g330399; PIDN:AAA45889.1; PID:g555157
C;Superfamily: Epstein-Barr virus nuclear antigen
C;Keywords: DNA binding; transcription regulation

  Query Match             9.9%;  Score 214;  DB 1;  Length 641;
  Best Local Similarity   39.7%;  Pred. No. 3.7e-06;
  Matches   75;  Conservative    1;  Mismatches   57;  Indels   56;  Gaps   11;

Qy       3 AHPGGGRCCPEQEEGESAAGGSGAGGDSAIEQGGQGSALAPSPVSGVRREGARGGGRGRG 62
           |  |||         |   |||:||||      || | | |     |    ||  || |||
Db     278 AGAGGG----AGGAGAGGAGGAGAGGAGGAGAGGAGGAGAG---GGAGAGGAGAGGGGRG 330

Qy      63 RWKQAGRG-GGVCGRGRG-----RGRG----------RGRGRGRGRGRGRPPSGGSGLGG 106
           |    ||| ||  |||||     ||||          | ||||||||  || |  |
Db     331 RGGSGGRGRGGSGGRGRGGSGGRRGRGRERARGGSRERARGRGRGRGEKRPRSPSS---- 386

Qy     107 DGGGCGGGGSGGGGAPRREPV---PF--PSGSAG-------PGPRG-----PRATESGKR 149
                 |  |  ||| |  .  ||  | | |          || |     | | | |
Db     387 -------QSSSSGSPPRRPPPGRRPFFHPVGEADYFEYHQEGGPDGEPDVPPGAIEQG-- 437

Qy     150 MDCPALPPG 158
           ||  ||
Db     438 ---PADDPG 443


RESULT 8
S31415
glycine-rich protein GRP22 - rape
C;Species: Brassica napus (rape)
```

```
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: S31415
R;Bergeron, D.; Boivin, R.; Baszczynski, C.L.; Bellemare, G.
submitted to the EMBL Data Library, August 1992
A;Description: Characterization and expression of a gene family encoding glycine-rich pr
A;Reference number: S31415
A;Accession: S31415
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-291 <BER>
A;Cross-references: EMBL:Z15045; NID:g17820; PIDN:CAA78762.1; PID:g17821
C;Superfamily: Phaseolus glycine-rich cell wall protein 1.8

  Query Match             9.8%;  Score 211.5;  DB 1;  Length 291;
  Best Local Similarity   39.4%;  Pred. No. 2.1e-06;
  Matches   61;  Conservative    1;  Mismatches   58;  Indels   35;  Gaps    5;

Qy       6 GGGRCCPEQEEGESAAGGSGAGGDSAIEQGGQGSALAPSPVSGVRREGARGGGRGRGRWK 65
           |||        ||  |  || | ||      || |          |    || ||| | |
Db     145 GGG-------EGAGAGGGYGGGG-----AGGHGGGGGGGNGGGGGGGGAHGGGYGGGEGA 192

Qy      66 QAG---RGGGVCGRGRGRGRGRGRGRGRGRGRPPSGGSGLGGDG----------GGCG 112
             ||    |||  | | | | |:| | | | | |    || | ||           || |
Db     193 GAGGGYGGGGAGGHGGGGGGGKGGGGGGGSGAGGAHGGGYGAGGGAGEGYGGGGGEGGHG 252

Qy     113 GGGSGGGGAPRREPVPFPSGSAGPGPRGPRATESG 147
           ||| |||||          |  | |  |  |  ||
Db     253 GGGGGGGGA----------GGGGGGGGGYAAAGSG 277


RESULT 9
T49109
glycine-rich protein - Arabidopsis thaliana
N;Alternate names: protein AT4g22020
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Sep-2000
C;Accession: T49109
R;Bevan, M.; Wedler, H.; Wambutt, R.; Bancroft, I.; Mewes, H.W.; Rudd, S.; Lemcke, K.; M
submitted to the Protein Sequence Database, May 2000
A;Reference number: Z25016
A;Accession: T49109
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-396 <BEV>
A;Cross-references: EMBL:AL022140; GSPDB:GN00062; ATSP:AT4g22020
A;Experimental source: cultivar Columbia; BAC clone F1N20
C;Genetics:
A;Gene: ATSP:AT4g22020
A;Map position: 4
C;Superfamily: Phaseolus glycine-rich cell wall protein 1.8

  Query Match             9.6%;  Score 208.5;  DB 2;  Length 396;
  Best Local Similarity   34.4%;  Pred. No. 4.6e-06;
  Matches   55;  Conservative    3;  Mismatches   63;  Indels   39;  Gaps    3;

Qy       6 GGGRCCPEQEEGESAAGGSGAGGDSAIEQGGQGSALAPSPVSGVRREGARGGGRGRGRWK 65
           |||        |    || | |||      | |:    |   ||   | |||| | |
Db      62 GGG--------GGGGGGGGGEGGDGYGHGEGYGAGAGMSGYGGVGGGGGRGGGEGGGSSA 113

Qy      66 QAGRG------------------GGVCGRGRGRGRGRGRGRGRGRGRPPSGGSGLGGD 107
             | |                  ||  | | | | | |   | | | |      |:|:||
Db     114 NGGSGHGSGSGAGAGVGGTTGGVGGGGGGGGGGGEGGGSSGGSGHGSGSGAGAGAGVGGS 173

Qy     108 GGGCGGGGSGGGGAPRREPVPFPSGSAGPGPRGPRATESG 147
           || |||| ||||              | |  |     ||
Db     174 SGGAGGGGGGGGG------------EGGGANGGSGHGSG 200


RESULT 10
S01820
glycine-rich cell wall protein 1.8 precursor - kidney bean
C;Species: Phaseolus vulgaris (kidney bean)
C;Date: 30-Sep-1989 #sequence_revision 19-May-1994 #text_change 16-Jul-1999
C;Accession: S01820
R;Keller, B.; Sauer, N.; Lamb, C.J.
EMBO J. 7, 3625-3633, 1988
A;Title: Glycine-rich cell wall proteins in bean: gene structure and association of t
A;Reference number: S01820; MUID:89091109; PMID:3208742
A;Accession: S01820
A;Molecule type: DNA
A;Residues: 1-465 <KEL>
A;Cross-references: EMBL:X13596; NID:g21002; PIDN:CAA31932.1; PID:g21003
C;Comment: This protein is enriched in the cell wall fraction of young hypocotyls and
tyls.
C;Comment: Much of the sequence consists of tandemly repeated 22-residue segments wit
C;Superfamily: Phaseolus glycine-rich cell wall protein 1.8
C;Keywords: cell wall; structural protein; tandem repeat
F;1-30/Domain: signal sequence #status predicted <SIG>
F;31-465/Product: glycine-rich cell wall protein 1.8 #status predicted <MAT>

  Query Match             9.6%;  Score 208.5;  DB 1;  Length 465;
  Best Local Similarity   42.1%;  Pred. No. 5.7e-06;
  Matches   59;  Conservative    5;  Mismatches   53;  Indels   23;  Gaps    6;

Qy       6 GGGRCCPEQEEGESAAGGSGAGGDSAIEQGGQGSALAPSPVSGVRREGARGGGRGRGRWK 65
           |||        :|  | || ||||:  |  || | : |      |    ||:||| | |
Db     120 GGG-------QGGGAGGGYGAGGEHGIGYGGGGGSGAGGG-GGYNAGGAQGGGYGTGGGA 171

Qy      66 -QAGRGGGVCGRGRGRGRGRGRGRGRGRGRPPSGGSGLGGDGGGCGG----GGSGGGG 120
            | |||  | | | |:| | | | | | |    || | || ||| ||    ||  |||
Db     172 GGGGGGGGDHGGGYGGGQGAGGGAGGGYGGGGEHGGGGG-GGQGGGAGGGYGAGGEHGGG 230

Qy     121 APRREPVPFPSGSAGPGPRG 140
           |          |  | |  |
Db     231 A---------GGGQGGGAGG 241


RESULT 11
B84777
hypothetical protein At2g36120 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C;Accession: B84777
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: B84777
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-255 <STO>
A;Cross-references: GB:AE002093; NID:g4678224; PIDN:AAD26969.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g36120
A;Map position: 2
C;Superfamily: collagen alpha 1(V) chain; fibrillar collagen carboxyl-terminal homolo

  Query Match             9.6%;  Score 207;  DB 2;  Length 255;
  Best Local Similarity   41.0%;  Pred. No. 3.4e-06;
  Matches   59;  Conservative    2;  Mismatches   61;  Indels   22;  Gaps    5;

Qy       5 PGGGRCCPEQEEGESAAGGSGAGGDSAIEQGGQGSALAPSPVSGVRREGARGGGRGRGRW 64
           ||||        |    ||:| ||  |   || |   |      |    || ||| | |:
Db     100 PGGG-----YGGGSGEGGGAGYGGGEAGGHGGGGGGGAGG--GGGGGGGAHGGGYGGGQG 152

Qy      65 KQAG---RGGGVCGRGRGRGRGRGRGRGRGRPPSGGSGLGGD-----GGGCGGGGS 116
             ||    |||  | | | | | | | | | | | |    || | ||      ||| | ||
Db     153 AGAGGGYGGGGAGGHGGGGGGGNGGGGGGGSGEGGAHGGGYGAGGGAGEGYGGGAGAGGH 212

Qy     117 GGGGAPRREPVPFPSGSAGPGPRG 140
           ||||          || | |  |
Db     213 GGGGGG-------GGGSGGGGGGG 229
```

```
RESULT 12
KNRZG2
glycine-rich cell wall structural protein 2 precursor - rice
C;Species: Oryza sativa (rice)
C;Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 16-Jul-1999
C;Accession: S18567
R;Fang, R.X.; Pang, Z.; Gao, D.M.; Mang, K.Q.; Chua, N.H.
Plant Mol. Biol. 17, 1255-1257, 1991
A;Title: cDNA sequence of a virus-inducible, glycine-rich protein gene from rice.
A;Reference number: S18567; MUID:92032791; PMID:1840687
A;Accession: S18567
A;Molecule type: mRNA
A;Residues: 1-183 <FAN>
A;Cross-references: EMBL:X54449; NID:g20244; PIDN:CAA38315.1; PID:g20245
C;Superfamily: glycine-rich cell wall structural protein 1
C;Keywords: cell wall; structural protein
F;1-27/Domain: signal sequence #status predicted <SIG>
F;28-183/Product: glycine-rich cell wall structural protein 2 #status predicted <MAT>

  Query Match             9.5%;  Score 206.5;  DB 1;  Length 183;
  Best Local Similarity   41.2%;  Pred. No. 2.4e-06;
  Matches   54;  Conservative    5;  Mismatches   53;  Indels   19;  Gaps    3;

Qy       5 PGGGRCCPEQEEGESAAGGSGAGGDSAIEQGGQGSALAPSPVSGVRREGARGGGRGRGRW 64
           ||||              || | ||      | ||        ||    || ||| |||
Db      31 PGGG----------GGGGGEGGGGGYGGSGYGSGSGYGEGGGSG----GAAGGGYGRG-- 74

Qy      65 KQAGRGGGVCGRGRGRGRGRGRGRGRGRGRPPSGGSGLGGDGGGCGGGGSGGGGAPRR 124
             | |||  | | | | | | |:| | | |   :|| | || |||  |||:|| |
Db      75 ---GGGGGGGGEGGGSGSGYGSGQGSGYGAGVGGAGGYGSGGGGGGGQGGGAGGYGQGSG 131

Qy     125 EPVPFPSGSAG 135
               : ||: |
Db     132 YGSGYGSGAGG 142


RESULT 13
T31611
hypothetical protein Y50E8A.g - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T31611
R;Steward, C.
submitted to the EMBL Data Library, September 1999
A;Reference number: Z21047
A;Accession: T31611
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1585 <WIL>
A;Cross-references: EMBL:AL117200; NID:e1549770; PIDN:CAB55050.1; CESP:Y50E8A.g
A;Experimental source: clone Y50E8A
C;Genetics:
A;Gene: CESP:Y50E8A.g
A;Introns: 25/3; 60/1; 133/2; 217/3; 270/3; 337/2; 400/1; 746/2

  Query Match             9.4%;  Score 204.5;  DB 2;  Length 1585;
  Best Local Similarity   35.8%;  Pred. No. 4.5e-05;
  Matches   62;  Conservative   11;  Mismatches   63;  Indels   37;  Gaps    7;

Qy       6 GGGRCCPEQEEGESAAGG--SGAGGDSAIEQGGQGSALAPSPVSGVRREGARGGGRGRGR 63
           |||        | |: ||  || ||      || | |       |   : : ||| | |
Db     439 GGG--------GGSSGGGYASGGGGGGGYASGGGGGA------GGGYAKPSGGGGGGGGY 484

Qy      64 WKQAGRGGGVCGRGRGRGRGRGRGRGRGRPPSGGSGLGG----DGGGCGGGGSGGG 119
            | :| |||    | |   | |  :  | | |    || |  |     ||| ||| |||
Db     485 AKPSGGGGGYASGGGGGSSGGGYAKSSGGGGGYASGGGGGSSGGSYASGGGGGGGSSGGS 544

Qy     120 G-------APRREPVPFPSGSAGPGPRGPRATESGKRMDCPALPPGWKKEEVI 165
           |       ||   | | |: :  | | |  |  ||:  |        ::|||:
Db     545 GGGYSSAAAPPPPPPPPPAPAPAPAPSGGYA--SGEVED--------RQEEVM 587


RESULT 14
C84470
hypothetical protein At2g05580 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 01-Mar-2002
C;Accession: C84470
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: C84470
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-302 <STO>
A;Cross-references: GB:AE002093; NID:g4581166; PIDN:AAD24649.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g05580
A;Map position: 2
C;Superfamily: collagen alpha 1(XIV) chain; fibronectin type III repeat homology; von

  Query Match             9.3%;  Score 202.5;  DB 2;  Length 302;
  Best Local Similarity   40.0%;  Pred. No. 8e-06;
  Matches   54;  Conservative    9;  Mismatches   53;  Indels   19;  Gaps    4;

Qy       6 GGGRCCPEQEEGESAAGGSGAGGDSAIEQGGQGSALAPSPVSGVRREGARGGGRGRGRWK 65
             |||      : |:   || | ||      ||||          |  : | :||| | |:
Db     183 GGG------QGGQKGGGGRGQGGMKGGGGGGQGGHKGGG--GGGGQGGHKGGGGGGGQGG 234

Qy      66 QAGRGGGVCGRGRGRGRGRGRGRGRGRGRPPSGGSGLGGDGGGCGGGGSGGGGAPRRE 125
             | |||  | |  :| | |:|  :| | |    || | || ||| |||||||||
Db     235 HKGGGGGGQGGGGHKGGGGGQGGHKG-GGGGGHVGGGGRGGGGGGRGGGGSGGGG----- 288

Qy     126 PVPFPSGSAGPGPRG 140
                | :| |  |
Db     289 -----GGGSGRGGGG 298


RESULT 15
KNRZG1
glycine-rich cell wall structural protein 1 precursor (clone lambda-313) - rice
C;Species: Oryza sativa (rice)
C;Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 16-Jul-1999
C;Accession: S13385
R;Lei, M.; Wu, R.
Plant Mol. Biol. 16, 187-198, 1991
A;Title: A novel glycine-rich cell wall protein gene in rice.
A;Reference number: S13385; MUID:91370862; PMID:1716496
A;Accession: S13385
A;Molecule type: DNA
A;Residues: 1-165 <LEI>
A;Cross-references: EMBL:X53596; NID:g20246; PIDN:CAA37665.1; PID:g20247
C;Genetics:
A;Gene: grp-1
C;Superfamily: glycine-rich cell wall structural protein 1
C;Keywords: cell wall; duplication; structural protein
F;1-23/Domain: signal sequence #status predicted <SIG>
F;24-165/Product: glycine-rich cell wall structural protein 1 #status predicted <MAT>
F;30-55/Region: repeat R1
F;56-62/Region: repeat R2
F;62-92/Region: repeat R1
F;93-99/Region: repeat R2
F;100-131/Region: repeat R1
F;132-138/Region: repeat R2
F;139-160/Region: repeat R1

  Query Match             9.3%;  Score 200.5;  DB 1;  Length 165;
  Best Local Similarity   40.8%;  Pred. No. 5.2e-06;
  Matches   51;  Conservative    6;  Mismatches   59;  Indels    9;  Gaps    3;

Qy      17 GESAAGGSGAGGDSAIEQGGQGSALAPSPVSGVRREGAR-GGGRGRGRWKQAGRGGGVCG 75
           | |  || | ||      || ||        |  : | : |||:| |     | |||  |
Db      32 GGSGGGGGGGGGGGG---GGNGSGSGSGYGYGYGKAGGQSGGGQGSG-----GGGGGGGG 83
```

```
Qy      76 RGRGRGRGRGRGRGRGRGRPPSGGSGLGGDGGGCGGGGSGGGGAPRREPVPFPSGSAG 135
           | | | | | | | |:| |     ||| || ||| ||||  | |:       :  | |
Db      84 GGNGSGSGSGYGYGYGQGNGGAQGQGSGGGGGGGGGGGGGGSGQGSGSGYGYGYGKGGGG 143

Qy     136 PGPRG 140
           |  |
Db     144 GGGGG 148
```

Search completed: March 12, 2003, 09:13:50
Job time : 39.9242 secs

GenCore version 5.1.4_p5_4578

OM protein - protein search, using sw model

Run on: March 12, 2003, 05:39:48 ; Search time 18.9316 Seconds
(without alignments)
915.501 Million cell updates/sec

Title: US-09-554-414B-2
Perfect score: 2167
Sequence: 1 MRAHPGGGRCCPEQEEGESA.....LSRAADTEEMDIEMDSGDEA 411

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 188354 seqs, 42170167 residues

Total number of hits satisfying chosen parameters: 188354

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2167	100.0	411	10	US-09-749-728B-1	Sequence 1, Appli
2	209.5	9.7	201	10	US-09-848-990-22	Sequence 22, Appl
3	209.5	9.7	201	10	US-09-760-364-14	Sequence 14, Appl
4	203.5	9.4	357	10	US-09-864-761-35807	Sequence 35807, A
5	200.5	9.3	200	9	US-10-160-354-4	Sequence 4, Appli
6	200.5	9.3	200	9	US-09-990-940-21	Sequence 21, Appl
7	200.5	9.3	200	9	US-10-026-021-8	Sequence 8, Appli
8	200.5	9.3	200	9	US-10-161-165-3	Sequence 3, Appli
9	200.5	9.3	200	9	US-10-160-663-3	Sequence 3, Appli
10	200.5	9.3	200	10	US-09-798-584-18	Sequence 18, Appl
11	200.5	9.3	200	10	US-09-967-624-19	Sequence 19, Appl
12	200.5	9.3	200	10	US-09-998-667-18	Sequence 18, Appl
13	189.5	8.7	651	10	US-09-861-597-1	Sequence 1, Appli
14	187.5	8.7	1079	9	US-09-820-843A-20	Sequence 20, Appl
15	183.5	8.5	283	10	US-09-864-761-36720	Sequence 36720, A
16	182.5	8.4	595	9	US-09-854-133-187	Sequence 187, App
17	182.5	8.4	595	10	US-09-738-973-187	Sequence 187, App
18	182.5	8.4	606	10	US-09-861-597-4	Sequence 4, Appli
19	180.5	8.3	484	9	US-09-820-843A-19	Sequence 19, Appl
20	178	8.2	645	10	US-09-919-172-41	Sequence 41, Appl
21	176.5	8.1	606	10	US-09-861-597-6	Sequence 6, Appli
22	175.5	8.1	606	10	US-09-861-597-8	Sequence 8, Appli
23	174.5	8.1	334	10	US-09-925-301-1363	Sequence 1363, Ap
24	170.5	7.9	256	9	US-09-820-843A-18	Sequence 18, Appl
25	169	7.8	1004	9	US-09-738-626-5676	Sequence 5676, Ap
26	162	7.5	218	10	US-09-925-300-1671	Sequence 1671, Ap
27	159.5	7.4	618	10	US-09-925-300-1381	Sequence 1381, Ap
28	156	7.2	141	10	US-09-864-761-36181	Sequence 36181, A
29	155	7.2	2211	9	US-10-096-961-1	Sequence 1, Appli
30	151	7.0	714	10	US-09-978-242-3	Sequence 3, Appli
31	150.5	6.9	101	10	US-09-861-597-3	Sequence 3, Appli
32	150	6.9	166	10	US-09-837-969A-21	Sequence 21, Appl
33	150	6.9	166	10	US-09-841-321A-21	Sequence 21, Appl
34	147	6.8	173	10	US-09-864-761-36371	Sequence 36371, A
35	146	6.7	191	10	US-09-864-761-36985	Sequence 36985, A
36	145.5	6.7	378	10	US-09-849-967A-2	Sequence 2, Appli
37	144.5	6.7	440	9	US-10-066-500-106	Sequence 106, App
38	144.5	6.7	440	9	US-10-063-547-52	Sequence 52, Appl
39	144.5	6.7	440	9	US-10-174-590-202	Sequence 202, App
40	144.5	6.7	440	9	US-10-176-758-202	Sequence 202, App
41	144.5	6.7	440	9	US-10-063-616-52	Sequence 52, Appl
42	144.5	6.7	440	9	US-10-175-737-202	Sequence 202, App
43	144.5	6.7	440	9	US-10-063-502-52	Sequence 52, Appl
44	144.5	6.7	440	9	US-10-173-706-202	Sequence 202, App
45	144.5	6.7	440	9	US-10-175-738-202	Sequence 202, App

ALIGNMENTS

```
RESULT 1
US-09-749-728B-1
; Sequence 1, Application US/09749728B
; Patent No. US20020142457A1
; GENERAL INFORMATION:
;  APPLICANT: Umezawa, Akihiro
;  APPLICANT:  Hata, Jun-Ichi
;  APPLICANT:  Fukuda, Keiichi
;  APPLICANT:  Ogawa, Satoshi
;  APPLICANT:  Sakurada, Kazuhiro
;  APPLICANT:  Gojo, Satoshi
;  APPLICANT:  Yamada, Yoji
;  TITLE OF INVENTION: THE CELL HAVING THE POTENTIALITY OF DIFFERENTIATION INTO CARDI
;  FILE REFERENCE: 00766.000043
;  CURRENT APPLICATION NUMBER: US/09/749,728B
;  CURRENT FILING DATE:  2001-09-17
;  PRIOR APPLICATION NUMBER: H11-372826
;  PRIOR FILING DATE: 1999-12-28
;  PRIOR APPLICATION NUMBER: PCT-JP00-01148
;  PRIOR FILING DATE: 2000-02-28
;  PRIOR APPLICATION NUMBER: PCT-JP00-07741
;  PRIOR FILING DATE: 2000-11-02
;  NUMBER OF SEQ ID NOS: 80
;  SOFTWARE: PatentIn Ver.2.0
; SEQ ID NO 1
;   LENGTH: 411
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-749-728B-1

  Query Match           100.0%;  Score 2167;  DB 10;  Length 411;
  Best Local Similarity  100.0%;  Pred. No. 3.2e-144;
  Matches  411;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MRAHPGGGRCCPEQEEGESAAGGSGAGGDSAIEQGGQGSALAPSPVSGVRREGARGGGRG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MRAHPGGGRCCPEQEEGESAAGGSGAGGDSAIEQGGQGSALAPSPVSGVRREGARGGGRG 60

Qy      61 RGRWKQAGRGGGVCGRGRGRGRGRGRGRGRGRGRPPSGGSGLGGDGGGCGGGGSGGGG 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 RGRWKQAGRGGGVCGRGRGRGRGRGRGRGRGRGRPPSGGSGLGGDGGGCGGGGSGGGG 120
```

```
Qy    121 APRREPVPFPSGSAGPGPRGPRATESGKRMDCPALPPGWKKEEVIRKSGLSAGKSDVYYF 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 APRREPVPFPSGSAGPGPRGPRATESGKRMDCPALPPGWKKEEVIRKSGLSAGKSDVYYF 180

Qy    181 SPSGKKFRSKPQLARYLGNTVDLSSFDFRTGKMMPSKLQKNKQRLRNDPLNQNKGKPDLN 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 SPSGKKFRSKPQLARYLGNTVDLSSFDFRTGKMMPSKLQKNKQRLRNDPLNQNKGKPDLN 240

Qy    241 TTLPIRQTASIFKQPVTKVTNHPSNKVKSDPQRMNEQPRQLFWEKRLQGLSASDVTEQII 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TTLPIRQTASIFKQPVTKVTNHPSNKVKSDPQRMNEQPRQLFWEKRLQGLSASDVTEQII 300

Qy    301 KTMELPKGLQGVGPGSNDETLLSAVASALHTSSAPITGQVSAAVEKNPAVWLNTSQPLCK 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 KTMELPKGLQGVGPGSNDETLLSAVASALHTSSAPITGQVSAAVEKNPAVWLNTSQPLCK 360

Qy    361 AFIVTDEDIRKQEERVQQVRKKLEEALMADILSRAADTEEMDIEMDSGDEA 411
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 AFIVTDEDIRKQEERVQQVRKKLEEALMADILSRAADTEEMDIEMDSGDEA 411
```

RESULT 2
US-09-848-990-22
; Sequence 22, Application US/09848990
; Patent No. US20020048572A1
; GENERAL INFORMATION:
; APPLICANT: Shan, Bei
; APPLICANT: Schultz, Joshua
; APPLICANT: Tu, Hua
; APPLICANT: Tularik Inc.
; TITLE OF INVENTION: Treatment of Hypertriglyceridemia and Other Conditions
; TITLE OF INVENTION: Using LXR Modulators
; FILE REFERENCE: 018781-004910US
; CURRENT APPLICATION NUMBER: US/09/848,990
; CURRENT FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: US 60/201,601
; PRIOR FILING DATE: 2000-05-03
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 201
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:flexible linker
; NAME/KEY: MOD_RES
; LOCATION: (1)..(97)
; OTHER INFORMATION: Gly at positions 1-97 may be present or absent
; NAME/KEY: MOD_RES
; LOCATION: (105)..(201)
; OTHER INFORMATION: Gly at positions 105-201 may be present or absent
US-09-848-990-22

Query Match 9.7%; Score 209.5; DB 10; Length 201;
Best Local Similarity 40.7%; Pred. No. 9.2e-08;
Matches 55; Conservative 0; Mismatches 63; Indels 17; Gaps 2;

```
Qy      6 GGGRCCPEQEEGESAAGGSGAGGDSAIEQGGQGSALAPSPVSGVRREGARGGGRGRGRWK 65
            |||        |    || | ||      || |          |    |  ||| | |
Db     17 GGG--------GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 68

Qy     66 QAGRGGGVCGRGRGRGRGRGRGRGRGRGRPPSGGSGLGGDGGGCGGGGSGGGGAPRRE 125
           | |||  | | | | | | | | | | |  | || | || ||| |||| ||||
Db     69 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGPGGGGGGGGGGGGGGGGGGGGGG---- 124

Qy    126 PVPFPSGSAGPGPRG 140
                |  | |  |
Db    125 -----GGGGGGGGGG 134
```

RESULT 3
US-09-760-364-14
; Sequence 14, Application US/09760364
; Patent No. US20020152479A1
; GENERAL INFORMATION:
; APPLICANT: Lehmann, Juergen Michael
; APPLICANT: Shiau, Andrew Kwan-Nan
; APPLICANT: Tularik Inc.
; TITLE OF INVENTION: CAR Modulators: Screening and Treatment of
; TITLE OF INVENTION: Hypercholesterolemia
; FILE REFERENCE: 018781-004110US
; CURRENT APPLICATION NUMBER: US/09/760,364
; CURRENT FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: US 60/176,398
; PRIOR FILING DATE: 2000-01-13
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 201
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:flexible linker
; NAME/KEY: MOD_RES
; LOCATION: (1)..(97)
; OTHER INFORMATION: Gly residues from positions 1-97 may be present or
; OTHER INFORMATION: absent
; NAME/KEY: MOD_RES
; LOCATION: (105)..(201)
; OTHER INFORMATION: Gly residues from, positions 105-201 may be
; OTHER INFORMATION: present or absent
US-09-760-364-14

Query Match 9.7%; Score 209.5; DB 10; Length 201;
Best Local Similarity 40.7%; Pred. No. 9.2e-08;
Matches 55; Conservative 0; Mismatches 63; Indels 17; Gaps 2;

```
Qy      6 GGGRCCPEQEEGESAAGGSGAGGDSAIEQGGQGSALAPSPVSGVRREGARGGGRGRGRWK 65
            |||        |    || | ||      || |          |    |  ||| | |
Db     17 GGG--------GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 68

Qy     66 QAGRGGGVCGRGRGRGRGRGRGRGRGRGRPPSGGSGLGGDGGGCGGGGSGGGGAPRRE 125
           | |||  | | | | | | | | | | |  | || | || ||| |||| ||||
Db     69 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGPGGGGGGGGGGGGGGGGGGGGGG---- 124

Qy    126 PVPFPSGSAGPGPRG 140
                |  | |  |
Db    125 -----GGGGGGGGGG 134
```

RESULT 4
US-09-864-761-35807
; Sequence 35807, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30

```
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 35807
;   LENGTH: 357
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AL078472.1
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 27
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 19
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 34
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 7.1
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 28
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 43
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 20
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 25
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 18
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 21
US-09-864-761-35807

  Query Match             9.4%;  Score 203.5;  DB 10;  Length 357;
  Best Local Similarity   42.2%;  Pred. No. 4.8e-07;
  Matches   57;  Conservative    3;  Mismatches   46;  Indels   29;  Gaps    5;

Qy       6 GGGRCCPEQEEGESAAGGSGAGGDSAIEQGGQGSALAPSPVSGVRREGARGGGRGRGRWK 65
           |||        :| |  |||| ||      || ||       ||   :|  |||||
Db     128 GGGGDGGGAGDGGSGGGGSGDGG------GGGGSG------SGDGDDGGGGGGRG----- 170

Qy      66 QAGRGGGVCGRGRGRGRGRGRGRGRGRGRPPSGGSGLGGDGGGCGGGGSGGGGAPRRE 125
            :| |||  | | | |   | | | | | |    || | |||||| || | ||||
Db     171 FSGDGGG--GGGGGSGGDAGGGGGSGDGGGGDGGGGDGGGGDGGGGGGSGDGGGG----- 223

Qy     126 PVPFPSGSAGPGPRG 140
               |  |  |  |
Db     224 -----DGGGGDGGGG 233


RESULT 5
US-10-160-354-4
; Sequence 4, Application US/10160354
; Publication No. US20030013107A1
; GENERAL INFORMATION:
;  APPLICANT: Chu, Peter
;  APPLICANT:  Li, Congfen
;  APPLICANT:  Liao, X. Charlene
;  APPLICANT:  Pardo, Jorge
;  APPLICANT:  Rigel Pharmaceuticals, Incorporated
;  TITLE OF INVENTION: Alpha 2 Intergrin:  Modulators of Lymphocyte Activation
;  FILE REFERENCE: 021044-001110US
;  CURRENT APPLICATION NUMBER: US/10/160,354
;  CURRENT FILING DATE:  2002-05-30
;  PRIOR APPLICATION NUMBER: US 60/296,819
;  PRIOR FILING DATE: 2001-06-07
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
;   LENGTH: 200
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:flexible linker
;   FEATURE:
;   NAME/KEY: MOD_RES
;   LOCATION: (6)..(200)
;   OTHER INFORMATION: Gly at positions 6-200 may be present or absent
US-10-160-354-4

  Query Match             9.3%;  Score 200.5;  DB 9;  Length 200;
  Best Local Similarity   40.0%;  Pred. No. 3.9e-07;
  Matches   54;  Conservative    0;  Mismatches   64;  Indels   17;  Gaps    2;

Qy       6 GGGRCCPEQEEGESAAGGSGAGGDSAIEQGGQGSALAPSPVSGVRREGARGGGRGRGRWK 65
           |||        |    || | ||      || |          |    |  ||| | |
Db       1 GGG--------GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 52

Qy      66 QAGRGGGVCGRGRGRGRGRGRGRGRGRGRPPSGGSGLGGDGGGCGGGGSGGGGAPRRE 125
             | |||  | | | | | | | | | | |    || | || ||| |||| ||||
Db      53 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG---- 108

Qy     126 PVPFPSGSAGPGPRG 140
               |  |  |  |
Db     109 -----GGGGGGGGGG 118


RESULT 6
US-09-990-940-21
; Sequence 21, Application US/09990940
; Publication No. US20030027252A1
; GENERAL INFORMATION:
;  APPLICANT: Tian, Hui
;  APPLICANT:  Zhao, Jiagang
;  APPLICANT:  Chen, Jin-Long
;  APPLICANT:  Cutler, Gene
;  APPLICANT:  An, Songzhu
;  APPLICANT:  Dai, Kang
;  APPLICANT:  Gupte, Jamila S.
;  APPLICANT:  Tularik Inc.
;  TITLE OF INVENTION: No. US20030027252A1el Receptors
;  FILE REFERENCE: 018781-007410US
;  CURRENT APPLICATION NUMBER: US/09/990,940
;  CURRENT FILING DATE:  2001-11-21
;  PRIOR APPLICATION NUMBER: US 60/252,841
;  PRIOR FILING DATE: 2000-11-22
;  PRIOR APPLICATION NUMBER: US 60/257,636
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: US 60/261,377
;  PRIOR FILING DATE: 2001-01-12
;  PRIOR APPLICATION NUMBER: US 60/279,554
;  PRIOR FILING DATE: 2001-03-28
;  PRIOR APPLICATION NUMBER: US 60/280,696
;  PRIOR FILING DATE: 2001-03-29
;  NUMBER OF SEQ ID NOS: 54
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
;   LENGTH: 200
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:flexible linker
;   NAME/KEY: MOD_RES
;   LOCATION: (6)..(200)
;   OTHER INFORMATION: Gly at positions 6-200 may be present or absent
;
```

US-09-990-940-21

Query Match 9.3%; Score 200.5; DB 9; Length 200;
Best Local Similarity 40.0%; Pred. No. 3.9e-07;
Matches 54; Conservative 0; Mismatches 64; Indels 17; Gaps 2;

```
Qy       6 GGGRCCPEQEEGESAAGGSGAGGDSAIEQGGQGSALAPSPVSGVRREGARGGGRGRGRWK 65
           |||         |    || | ||      || |          |    |  ||| | |
Db       1 GGG--------GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 52

Qy      66 QAGRGGGVCGRGRGRGRGRGRGRGRGRGRGRPPSGGSGLGGDGGGCGGGGSGGGGAPRRE 125
             | |||  | | | | | | | | | | |    || | || ||| |||| ||||
Db      53 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG---- 108

Qy     126 PVPFPSGSAGPGPRG 140
               |  | |  |
Db     109 -----GGGGGGGGG 118
```

RESULT 7
US-10-026-021-8
; Sequence 8, Application US/10026021
; Publication No. US20030027756A1
; GENERAL INFORMATION:
; APPLICANT: Hitoshi, Yasumichi
; APPLICANT: Demo, Susan
; APPLICANT: Jenkins, Yonchu
; APPLICANT: Rigel Pharmaceuticals, Inc.
; TITLE OF INVENTION: SAK: Modulation of Cellular Proliferation for
; TITLE OF INVENTION: Treatment of Cancer
; FILE REFERENCE: 021044-001210US
; CURRENT APPLICATION NUMBER: US/10/026,021
; CURRENT FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: US 60/309,632
; PRIOR FILING DATE: 2001-08-01
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 200
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:flexible linker
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (6)..(200)
; OTHER INFORMATION: Gly at positions 6-200 may be present or absent
US-10-026-021-8

Query Match 9.3%; Score 200.5; DB 9; Length 200;
Best Local Similarity 40.0%; Pred. No. 3.9e-07;
Matches 54; Conservative 0; Mismatches 64; Indels 17; Gaps 2;

```
Qy       6 GGGRCCPEQEEGESAAGGSGAGGDSAIEQGGQGSALAPSPVSGVRREGARGGGRGRGRWK 65
           |||         |    || | ||      || |          |    |  ||| | |
Db       1 GGG--------GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 52

Qy      66 QAGRGGGVCGRGRGRGRGRGRGRGRGRGRGRPPSGGSGLGGDGGGCGGGGSGGGGAPRRE 125
             | |||  | | | | | | | | | | |    || | || ||| |||| ||||
Db      53 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG---- 108

Qy     126 PVPFPSGSAGPGPRG 140
               |  | |  |
Db     109 -----GGGGGGGGG 118
```

RESULT 8
US-10-161-165-3
; Sequence 3, Application US/10161165
; Publication No. US20030027763A1
; GENERAL INFORMATION:
; APPLICANT: Bennett, Mark
; APPLICANT: Holland, Sacha
; APPLICANT: Rossi, Alex
; APPLICANT: Rigel Pharmaceuticals, Incorporated
; TITLE OF INVENTION: CD43: Modulators of Mast Cell Degranulation
; FILE REFERENCE: 021044-001010US
; CURRENT APPLICATION NUMBER: US/10/161,165
; CURRENT FILING DATE: 2002-05-31
; PRIOR APPLICATION NUMBER: US 60/296,801
; PRIOR FILING DATE: 2001-06-07
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 200
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:poly Gly tag
; OTHER INFORMATION: flexible linker
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (6)..(200)
; OTHER INFORMATION: Gly residues from position 6 to 200 may be present
; OTHER INFORMATION: or absent
US-10-161-165-3

Query Match 9.3%; Score 200.5; DB 9; Length 200;
Best Local Similarity 40.0%; Pred. No. 3.9e-07;
Matches 54; Conservative 0; Mismatches 64; Indels 17; Gaps 2;

```
Qy       6 GGGRCCPEQEEGESAAGGSGAGGDSAIEQGGQGSALAPSPVSGVRREGARGGGRGRGRWK 65
           |||         |    || | ||      || |          |    |  ||| | |
Db       1 GGG--------GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 52

Qy      66 QAGRGGGVCGRGRGRGRGRGRGRGRGRGRGRPPSGGSGLGGDGGGCGGGGSGGGGAPRRE 125
             | |||  | | | | | | | | | | |    || | || ||| |||| ||||
Db      53 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG---- 108

Qy     126 PVPFPSGSAGPGPRG 140
               |  | |  |
Db     109 -----GGGGGGGGG 118
```

RESULT 9
US-10-160-663-3
; Sequence 3, Application US/10160663
; Publication No. US20030040001A1
; GENERAL INFORMATION:
; APPLICANT: Demo, Susan
; APPLICANT: Hitoshi, Yasumichi
; APPLICANT: Pearsall, Denise
; APPLICANT: Rigel Pharmaceutical, Incorporated
; TITLE OF INVENTION: LETM1: Modulators of Cellular Proliferation
; FILE REFERENCE: 021044-000920US
; CURRENT APPLICATION NUMBER: US/10/160,663
; CURRENT FILING DATE: 2002-05-31
; PRIOR APPLICATION NUMBER: US 60/296,817
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: US 60/347,970
; PRIOR FILING DATE: 2002-01-15
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 200
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:poly Gly tag
; OTHER INFORMATION: flexible linker
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (6)..(200)
; OTHER INFORMATION: Gly residues from position 6 to 200 may be present
; OTHER INFORMATION: or absent
US-10-160-663-3

```
  Query Match             9.3%;  Score 200.5;  DB 9;  Length 200;
  Best Local Similarity   40.0%;  Pred. No. 3.9e-07;
  Matches   54;  Conservative    0;  Mismatches   64;  Indels   17;  Gaps    2;

Qy      6 GGGRCCPEQEEGESAAGGSGAGGDSAIEQGGQGSALAPSPVSGVRREGARGGGRGRGRWK 65
          |||         |    || | ||      || |          |    |  ||| | |
Db      1 GGG--------GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 52

Qy     66 QAGRGGGVCGRGRGRGRGRGRGRGRGRGRGRPPSGGSGLGGDGGGCGGGGSGGGGAPRRE 125
          | |||  | | | | | | | | | | |    || | || ||| |||| ||||
Db     53 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG---- 108

Qy    126 PVPFPSGSAGPGPRG 140
                | | | |
Db    109 -----GGGGGGGGG 118


RESULT 10
US-09-798-584-18
; Sequence 18, Application US/09798584
; Patent No. US20020102676A1
; GENERAL INFORMATION:
;  APPLICANT: Mu, David
;  APPLICANT:  Powers, Scott
;  APPLICANT:  Tularik Inc.
;  TITLE OF INVENTION: KCNB:  A No. US20020102676A1el Potassium Channel Protein
;  FILE REFERENCE: 018781-004010US
;  CURRENT APPLICATION NUMBER: US/09/798,584
;  CURRENT FILING DATE:  2001-03-03
;  PRIOR APPLICATION NUMBER: US 60/186,951
;  PRIOR FILING DATE: 2000-03-03
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
;   LENGTH: 200
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:poly-Gly linker
;   NAME/KEY: MOD_RES
;   LOCATION: (6)..(200)
;   OTHER INFORMATION: Gly at positions 6-200 may be present or absent
US-09-798-584-18

  Query Match             9.3%;  Score 200.5;  DB 10;  Length 200;
  Best Local Similarity   40.0%;  Pred. No. 3.9e-07;
  Matches   54;  Conservative    0;  Mismatches   64;  Indels   17;  Gaps    2;

Qy      6 GGGRCCPEQEEGESAAGGSGAGGDSAIEQGGQGSALAPSPVSGVRREGARGGGRGRGRWK 65
          |||         |    || | ||      || |          |    |  ||| | |
Db      1 GGG--------GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 52

Qy     66 QAGRGGGVCGRGRGRGRGRGRGRGRGRGRGRPPSGGSGLGGDGGGCGGGGSGGGGAPRRE 125
          | |||  | | | | | | | | | | |    || | || ||| |||| ||||
Db     53 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG---- 108

Qy    126 PVPFPSGSAGPGPRG 140
                | | | |
Db    109 -----GGGGGGGGG 118


RESULT 11
US-09-967-624-19
; Sequence 19, Application US/09967624
; Patent No. US20020142325A1
; GENERAL INFORMATION:
;  APPLICANT: Liao, X. Charlene
;  APPLICANT:  Chu, Peter
;  APPLICANT:  Pardo, Jorge
;  APPLICANT:  Li, Congfen
;  APPLICANT:  Zhao, Haoran
;  APPLICANT:  Wu, Jun
;  APPLICANT:  Rigel Pharmaceuticals, Inc.
;  TITLE OF INVENTION: PAK2:  Modulators of Lymphocyte Activation
;  FILE REFERENCE: 021044-000700US
;  CURRENT APPLICATION NUMBER: US/09/967,624
;  CURRENT FILING DATE:  2001-09-28
;  PRIOR APPLICATION NUMBER: US 60/280,647
;  PRIOR FILING DATE: 2001-03-30
;  NUMBER OF SEQ ID NOS: 19
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 19
;   LENGTH: 200
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:poly Gly tag
;   OTHER INFORMATION: flexible linker
;   FEATURE:
;   NAME/KEY: MOD_RES
;   LOCATION: (6)..(200)
;   OTHER INFORMATION: Gly residues from position 6 to 200 may be present
;   OTHER INFORMATION: or absent
US-09-967-624-19

  Query Match             9.3%;  Score 200.5;  DB 10;  Length 200;
  Best Local Similarity   40.0%;  Pred. No. 3.9e-07;
  Matches   54;  Conservative    0;  Mismatches   64;  Indels   17;  Gaps    2;

Qy      6 GGGRCCPEQEEGESAAGGSGAGGDSAIEQGGQGSALAPSPVSGVRREGARGGGRGRGRWK 65
          |||         |    || | ||      || |          |    |  ||| | |
Db      1 GGG--------GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 52

Qy     66 QAGRGGGVCGRGRGRGRGRGRGRGRGRGRGRPPSGGSGLGGDGGGCGGGGSGGGGAPRRE 125
          | |||  | | | | | | | | | | |    || | || ||| |||| ||||
Db     53 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG---- 108

Qy    126 PVPFPSGSAGPGPRG 140
                | | | |
Db    109 -----GGGGGGGGG 118


RESULT 12
US-09-998-667-18
; Sequence 18, Application US/09998667
; Patent No. US20020146747A1
; GENERAL INFORMATION:
;  APPLICANT: Masuda, Esteban
;  APPLICANT:  Liao, X. Charlene
;  APPLICANT:  Zhao, Haoran
;  APPLICANT:  Chu, Peter
;  APPLICANT:  Pardo, Jorge
;  APPLICANT:  Rigel Pharmaceuticals, Incorporated
;  TITLE OF INVENTION: TRAC1:  Modulators of Lymphocyte Activation
;  FILE REFERENCE: 021044-000600US
;  CURRENT APPLICATION NUMBER: US/09/998,667
;  CURRENT FILING DATE:  2001-12-03
;  PRIOR APPLICATION NUMBER: US 60/282,432
;  PRIOR FILING DATE: 2001-04-06
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
;   LENGTH: 200
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:flexible linker
;   FEATURE:
;   NAME/KEY: MOD_RES
;   LOCATION: (6)..(200)
;   OTHER INFORMATION: Gly at positions 6-200 may be present or absent
US-09-998-667-18

  Query Match             9.3%;  Score 200.5;  DB 10;  Length 200;
  Best Local Similarity   40.0%;  Pred. No. 3.9e-07;
```

```
  Matches   54;  Conservative    0;  Mismatches   64;  Indels   17;  Gaps    2;

Qy       6 GGGRCCPEQEEGESAAGGSGAGGDSAIEQGGQGSALAPSPVSGVRREGARGGGRGRGRWK 65
           |||         |    || | ||      || |         |    |  ||| | |
Db       1 GGG--------GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 52

Qy      66 QAGRGGGVCGRGRGRGRGRGRGRGRGRGRGRPPSGGSGLGGDGGGCGGGGSGGGGAPRRE 125
             | |||  | | | | | | | | | | |    || | || ||| |||| ||||
Db      53 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG---- 108

Qy     126 PVPFPSGSAGPGPRG 140
                 |  | |  |
Db     109 -----GGGGGGGGG 118


RESULT 13
US-09-861-597-1
; Sequence 1, Application US/09861597
; Patent No. US20020064539A1
; GENERAL INFORMATION:
;  APPLICANT: PHILLIPPE, Michel
;  APPLICANT:  GARSON, Jean-Claude
;  APPLICANT:  ARRAUDEAU, Jean-Pierre
;  TITLE OF INVENTION: COSMETIC OR DERMATOLOGICAL COMPOSITION CONTACTING AT
;  TITLE OF INVENTION:  LEAST ONE NATURAL OR RECOMBINANT SPIDER SILK OR AN
;  TITLE OF INVENTION:  ANALOG
;  FILE REFERENCE: 6388-0365-0
;  CURRENT APPLICATION NUMBER: US/09/861,597
;  CURRENT FILING DATE:  2001-05-22
;  PRIOR APPLICATION NUMBER: 09/247,806
;  PRIOR FILING DATE: 1999-02-11
;  PRIOR APPLICATION NUMBER: FR 98/01614
;  PRIOR FILING DATE: 1998-02-11
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 651
;   TYPE: PRT
;   ORGANISM: Nephila clavipes
US-09-861-597-1

  Query Match             8.7%;  Score 189.5;  DB 10;  Length 651;
  Best Local Similarity   31.4%;  Pred. No. 9.4e-06;
  Matches   59;  Conservative   19;  Mismatches   49;  Indels   61;  Gaps    8;

Qy       6 GGGRCCPEQEEGESAAGGSGAG-----GDSAIEQGGQGSALAPSPVSGVRREGARGGGRG 60
           | ||     :   :|||||:| |     |     :||||:  | :   |  :|| || | |
Db     458 GAGR---GGQGAAAAAGGAGQGGYGGLGSQGAGRGGQGAGAAAAAAVGAGQEGIRGQGAG 514

Qy      61 RGRW-----KQAGRG------------------GGVCGRGRGRGRG------------- 83
           :| :     : :|||                    ||: |:| |:| |
Db     515 QGGYGGLGSQGSGRGGLGGQGAGAAAAAAGGAGQGGGLGGQGAGQGAGAAAAAAGGVRQGG 574

Qy      84 ------RGRGR-GRGRGRGRPPSGGSGLGGDGG----GCGGGGSGGGGAPRREPVPFPSG 132
                :| || |:| |     :||:| || ||    | | || || ||          :|
Db     575 YGGLGSQGAGRGGQGAGAAAAAAGGAGQGGYGGLGGQGVGRGGLGGQGAGAA-----AAG 629

Qy     133 SAGPGPRG 140
             || |  |
Db     630 GAGQGGYG 637


RESULT 14
US-09-820-843A-20
; Sequence 20, Application US/09820843A
; Publication No. US20030039963A1
; GENERAL INFORMATION:
;  APPLICANT: Council of Scientific and Industrial Research
;  TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEI
;  TITLE OF INVENTION:  USEFUL AS ANTI-INFECTIVES
;  FILE REFERENCE: Q63915
;  CURRENT APPLICATION NUMBER: US/09/820,843A
;  CURRENT FILING DATE:  2001-03-30
;  NUMBER OF SEQ ID NOS: 118
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 20
;   LENGTH: 1079
;   TYPE: PRT
;   ORGANISM: M. tuberculosis
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: PE_PGRS
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: gi|2924449
US-09-820-843A-20

  Query Match             8.7%;  Score 187.5;  DB 9;  Length 1079;
  Best Local Similarity   36.3%;  Pred. No. 2.4e-05;
  Matches   53;  Conservative    7;  Mismatches   63;  Indels   23;  Gaps    6;

Qy       7 GGRCCPEQEEGESA-AGGSGAGGDSAIE--QGGQGSALAPSPVSGVRREGARGGGR--GR 61
           ||   |    |:   ||| ||||    :   || |        :|    |  || |  |
Db     581 GGGTVPAGSGGQGGNAGGGAGGQGGADGGSGGDGGD------AGTGGNGGNGGNRNSGN 634

Qy      62 GRWKQAGRGGGVCGRGRGRGRGRGRGRGRGRPPSGGSGLGGDGGGCGGGGSGGGGA 121
           |     | |||    | |   | | | | | |     :| :| ||:| | | ||:|| |
Db     635 GTGGAGGNGGGGANGGAGGAGGSGGGTG-GNGGAGGDAGDAGNGGNGNGTGNGGNGGNG- 692

Qy     122 PRREPVPFPSGSAGPGPRGPRATESG 147
                 | || |  |   | ||
Db     693 ----------GIAGMGGNGGAGTGSG 708


RESULT 15
US-09-864-761-36720
; Sequence 36720, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
```

```
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 36720
;   LENGTH: 283
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC006547.9
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 17
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 11
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 8.6
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 9
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 12
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 11
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 10
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 8.4
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 10
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 12
US-09-864-761-36720

  Query Match             8.5%;  Score 183.5;  DB 10;  Length 283;
  Best Local Similarity   37.0%;  Pred. No. 9.1e-06;
  Matches   54;  Conservative    5;  Mismatches   66;  Indels   21;  Gaps    4;

Qy       6 GGGRCCPEQEEGESAAGGSGAGGDSAIEQGGQGSALAPSPVSGVRREGARGGGRGRGRWK 65
           |||         | |  || | ||  :   || ||         |    |: ||| | |
Db       4 GGG--------GGSDGGGGGDGGGGSDGGGGGGS-------DGGGGGGSDGGGGGDGGGG 48

Qy      66 QAGRGGGVCGRGRGRGRGRGRGRGRGRGRGRPPSGGSGLGGDGGGCGGGGSGGGGAPRRE 125
             | |||    | | |   | | |   | |    || | | |||| ||||  |||    :
Db      49 SDGGGGGGGSDGGGGGSDGGGGGGGSDGGGSGSDGGGG-GSDGGGGGGGGDDGGGGGGSD 107

Qy     126 PVPFPSGSAGPGPRGPRATESGKRMD 151
               |  | |  |   :: |   |
Db     108 -----GGGGGGGSDGDGGSDGGGGSD 128


Search completed: March 12, 2003, 09:14:38
Job time : 25.9316 secs
```

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: March 14, 2003, 03:37:13 ; Search time 1339 Seconds
(without alignments)
96.762 Million cell updates/sec

Title: CGCGCGCG
Perfect score: 8
Sequence: 1 cgcgcgcg 8

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	8	100.0	19	17	AZ858877	AZ858877 2M0164D14
c	2	8	100.0	19	17	AZ858877	AZ858877 2M0164D14
	3	8	100.0	23	17	AZ316719	AZ316719 1M0035A01
c	4	8	100.0	23	17	AZ316719	AZ316719 1M0035A01
	5	8	100.0	25	17	BH852925	BH852925 SALK_0757
c	6	8	100.0	25	17	BH852925	BH852925 SALK_0757
	7	8	100.0	28	17	AZ606831	AZ606831 1M0428N24
c	8	8	100.0	28	17	AZ606831	AZ606831 1M0428N24
	9	8	100.0	28	17	AZ627049	AZ627049 1M0467E21
c	10	8	100.0	28	17	AZ627049	AZ627049 1M0467E21
	11	8	100.0	30	13	BI556227	BI556227 603237625
c	12	8	100.0	30	13	BI556227	BI556227 603237625
	13	8	100.0	38	12	BF302851	BF302851 602032746
c	14	8	100.0	38	12	BF302851	BF302851 602032746
	15	8	100.0	38	13	BM400755	BM400755 5009-0-78
c	16	8	100.0	38	13	BM400755	BM400755 5009-0-78
	17	8	100.0	39	12	BG538554	BG538554 602567279
c	18	8	100.0	39	12	BG538554	BG538554 602567279
	19	8	100.0	40	17	AZ431922	AZ431922 1M0217D21
c	20	8	100.0	40	17	AZ431922	AZ431922 1M0217D21
	21	8	100.0	40	17	BH796426	BH796426 1008094B1
c	22	8	100.0	40	17	BH796426	BH796426 1008094B1
	23	8	100.0	41	17	BH639962	BH639962 1008033A0
c	24	8	100.0	41	17	BH639962	BH639962 1008033A0
	25	8	100.0	43	13	BG915507	BG915507 602815734
c	26	8	100.0	43	13	BG915507	BG915507 602815734
	27	8	100.0	44	17	BH629051	BH629051 1007076B0
c	28	8	100.0	44	17	BH629051	BH629051 1007076B0
	29	8	100.0	45	17	AZ514564	AZ514564 1M0361N10
c	30	8	100.0	45	17	AZ514564	AZ514564 1M0361N10
	31	8	100.0	47	17	AZ400633	AZ400633 1M0167F06
c	32	8	100.0	47	17	AZ400633	AZ400633 1M0167F06
	33	8	100.0	47	17	AZ585161	AZ585161 1M0390G12
c	34	8	100.0	47	17	AZ585161	AZ585161 1M0390G12
	35	8	100.0	48	17	AZ764402	AZ764402 1M0560O18
c	36	8	100.0	48	17	AZ764402	AZ764402 1M0560O18
	37	8	100.0	49	9	AA948394	AA948394 on52b09.s
c	38	8	100.0	49	9	AA948394	AA948394 on52b09.s
	39	8	100.0	49	9	AI440059	AI440059 ti61f10.x
c	40	8	100.0	49	9	AI440059	AI440059 ti61f10.x
	41	8	100.0	49	9	AI565007	AI565007 tq53d09.x
c	42	8	100.0	49	9	AI565007	AI565007 tq53d09.x
	43	8	100.0	49	17	AZ456464	AZ456464 1M0259L12
c	44	8	100.0	49	17	AZ456464	AZ456464 1M0259L12
	45	8	100.0	49	17	AZ773338	AZ773338 1M0584L13

ALIGNMENTS

RESULT 1
AZ858877
LOCUS AZ858877 19 bp DNA linear GSS 21-FEB-2001
DEFINITION 2M0164D14F Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC2M0164D14 F, DNA sequence.
ACCESSION AZ858877
VERSION AZ858877.1 GI:13052498
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 19)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00

```
            Plate: 0164  row: D  column: 14
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 19.
FEATURES             Location/Qualifiers
     source          1. .19
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0164D14"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT        0 a      9 c     10 g      0 t
ORIGIN

  Query Match           100.0%;  Score 8;  DB 17;  Length 19;
  Best Local Similarity  100.0%;  Pred. No. 3.4e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db      10 CGCGCGCG 17


RESULT 2
AZ858877/c
LOCUS       AZ858877                 19 bp    DNA     linear   GSS 21-FEB-2001
DEFINITION  2M0164D14F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC2M0164D14 F, DNA sequence.
ACCESSION   AZ858877
VERSION     AZ858877.1  GI:13052498
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 19)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
```

```
            Plate: 0164  row: D  column: 14
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 19.
FEATURES             Location/Qualifiers
     source          1. .19
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0164D14"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT        0 a      9 c     10 g      0 t
ORIGIN

  Query Match           100.0%;  Score 8;  DB 17;  Length 19;
  Best Local Similarity  100.0%;  Pred. No. 3.4e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       9 CGCGCGCG 2


RESULT 3
AZ316719
LOCUS       AZ316719                 23 bp    DNA     linear   GSS 29-SEP-2000
DEFINITION  1M0035A01F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0035A01 F, DNA sequence.
ACCESSION   AZ316719
VERSION     AZ316719.1  GI:10364814
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 23)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
```

```
           Plate: 0035  row: A  column: 01
           Seq primer: CGTTGTAAAACGACGGCCAGT
           Class: plasmid ends
           High quality sequence stop: 23.
FEATURES             Location/Qualifiers
     source          1. .23
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0035A01"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT        0 a     12 c     11 g      0 t
ORIGIN

  Query Match             100.0%;  Score 8;  DB 17;  Length 23;
  Best Local Similarity   100.0%;  Pred. No. 3.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db      15 CGCGCGCG 22


RESULT 4
AZ316719/c
LOCUS       AZ316719                23 bp    DNA     linear   GSS 29-SEP-2000
DEFINITION  1M0035A01F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0035A01 F, DNA sequence.
ACCESSION   AZ316719
VERSION     AZ316719.1  GI:10364814
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 23)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
```

```
           Plate: 0035  row: A  column: 01
           Seq primer: CGTTGTAAAACGACGGCCAGT
           Class: plasmid ends
           High quality sequence stop: 23.
FEATURES             Location/Qualifiers
     source          1. .23
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0035A01"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT        0 a     12 c     11 g      0 t
ORIGIN

  Query Match             100.0%;  Score 8;  DB 17;  Length 23;
  Best Local Similarity   100.0%;  Pred. No. 3.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db      10 CGCGCGCG 3


RESULT 5
BH852925
LOCUS       BH852925                25 bp    DNA     linear   GSS 13-JUN-2002
DEFINITION  SALK_075769.37.15.x Arabidopsis thaliana TDNA insertion lines
            Arabidopsis thaliana genomic clone SALK_075769.37.15.x, DNA
            sequence.
ACCESSION   BH852925
VERSION     BH852925.1  GI:21423796
KEYWORDS    GSS.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 25)
  AUTHORS   Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab
            ,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P.
            , Zimmerman,J. and Ecker,J.R.
  TITLE     A Sequence-Indexed Library of Insertion Mutations in the
            Arabidopsis Genome
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Joseph R. Ecker
            Salk Institute Genomic Analysis Laboratory (SIGnAL)
            The Salk Institute for Biological Studies
            10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
            Tel: 858 453 4100 x1752
            Fax: 858 558 6379
            Email: ecker@salk.edu
            This is single pass sequence recovered from the left border of
```

```
            TDNA. This sequence lies within 300 bases of the 5' end of
            At1g56600.
            Class: TDNA tagged.
FEATURES              Location/Qualifiers
     source           1. .25
                      /organism="Arabidopsis thaliana"
                      /strain="Columbia 0"
                      /db_xref="taxon:3702"
                      /clone="SALK_075769.37.15.x"
                      /clone_lib="Arabidopsis thaliana TDNA insertion lines"
                      /note="PCR was performed on Arabidopsis thaliana lines
                      each of which contains  one or more TDNA insertion
                      elements. The resultant fragment for each  line was
                      directly sequenced to determine the genomic sequence at
                      the  site of insertion. Details of the protocols used can
                      be found at  http://signal.salk.edu/tdna_protocols.html"
BASE COUNT        5 a      6 c      9 g      5 t
ORIGIN

  Query Match             100.0%;  Score 8;  DB 17;  Length 25;
  Best Local Similarity   100.0%;  Pred. No. 3.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db      13 CGCGCGCG 20


RESULT 6
BH852925/c
LOCUS       BH852925                  25 bp    DNA     linear   GSS 13-JUN-2002
DEFINITION  SALK_075769.37.15.x Arabidopsis thaliana TDNA insertion lines
            Arabidopsis thaliana genomic clone SALK_075769.37.15.x, DNA
            sequence.
ACCESSION   BH852925
VERSION     BH852925.1  GI:21423796
KEYWORDS    GSS.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 25)
  AUTHORS   Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab
            ,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P.
            , Zimmerman,J. and Ecker,J.R.
  TITLE     A Sequence-Indexed Library of Insertion Mutations in the
            Arabidopsis Genome
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Joseph R. Ecker
            Salk Institute Genomic Analysis Laboratory (SIGnAL)
            The Salk Institute for Biological Studies
            10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
            Tel: 858 453 4100 x1752
            Fax: 858 558 6379
            Email: ecker@salk.edu
            This is single pass sequence recovered from the left border of
            TDNA. This sequence lies within 300 bases of the 5' end of
            At1g56600.
            Class: TDNA tagged.
FEATURES              Location/Qualifiers
     source           1. .25
                      /organism="Arabidopsis thaliana"
                      /strain="Columbia 0"
                      /db_xref="taxon:3702"
                      /clone="SALK_075769.37.15.x"
                      /clone_lib="Arabidopsis thaliana TDNA insertion lines"
                      /note="PCR was performed on Arabidopsis thaliana lines
                      each of which contains  one or more TDNA insertion
                      elements. The resultant fragment for each  line was
                      directly sequenced to determine the genomic sequence at
                      the  site of insertion. Details of the protocols used can
                      be found at  http://signal.salk.edu/tdna_protocols.html"
BASE COUNT        5 a      6 c      9 g      5 t
ORIGIN

  Query Match             100.0%;  Score 8;  DB 17;  Length 25;
  Best Local Similarity   100.0%;  Pred. No. 3.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db      20 CGCGCGCG 13


RESULT 7
AZ606831
LOCUS       AZ606831                  28 bp    DNA     linear   GSS 13-DEC-2000
DEFINITION  1M0428N24R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0428N24 R, DNA sequence.
ACCESSION   AZ606831
VERSION     AZ606831.1  GI:11729021
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 28)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0428  row: N  column: 24
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 28.
FEATURES              Location/Qualifiers
     source           1. .28
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="UUGC1M0428N24"
                      /clone_lib="Mouse 10kb plasmid UUGC1M library"
                      /sex="Male"
                      /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                      /note="Vector: PWD42nv; Purified genomic DNA from M.
                      musculus C57BL/6J (male) was obtained  from the Jackson
                      Laboratory Mouse DNA Resource
                      (http://www.jax.org/resources/documents/dnares/). The DNA
                      was hydrodynamically sheared by repeated passage through a
                      0.005 inch orifice at constant velocity. The sheared DNA
                      was blunt end-repaired with T4 DNA polymerase and T4
                      polynucleotide kinase. Adaptor oligonucleotides were
                      ligated to the blunt ends in high  molar excess. The
                      adaptored DNA was purified and size-selected for a 9.5 to
                      10.5 kb range using preparative agarose gel
                      electrophoresis. Vector DNA was prepared from a derivative
                      of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                      inducible derivative of plasmid R1. The vector was ligated
                      with adaptors complementary to the insert adaptors and
                      purified. The sheared, adaptored mouse DNA was annealed to
                      adaptored vector DNA, and transformed into
                      chemically-competent E. coli XL10-Gold (Stratagene) cells
                      and selected for ampicillin resistance."
```

```
BASE COUNT        0 a      13 c      15 g       0 t
ORIGIN

  Query Match           100.0%;  Score 8;  DB 17;  Length 28;
  Best Local Similarity  100.0%;  Pred. No. 3.2e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db     13 CGCGCGCG 20


RESULT 8
AZ606831/c
LOCUS       AZ606831                 28 bp    DNA     linear   GSS 13-DEC-2000
DEFINITION  1M0428N24R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0428N24 R, DNA sequence.
ACCESSION   AZ606831
VERSION     AZ606831.1  GI:11729021
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 28)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0428  row: N  column: 24
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 28.
FEATURES             Location/Qualifiers
     source          1. .28
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0428N24"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
```

```
BASE COUNT        0 a      13 c      15 g       0 t
ORIGIN

  Query Match           100.0%;  Score 8;  DB 17;  Length 28;
  Best Local Similarity  100.0%;  Pred. No. 3.2e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db     10 CGCGCGCG 3


RESULT 9
AZ627049
LOCUS       AZ627049                 28 bp    DNA     linear   GSS 13-DEC-2000
DEFINITION  1M0467E21R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0467E21 R, DNA sequence.
ACCESSION   AZ627049
VERSION     AZ627049.1  GI:11749239
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 28)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0467  row: E  column: 21
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 28.
FEATURES             Location/Qualifiers
     source          1. .28
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0467E21"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
```

```
BASE COUNT        0 a     14 c     14 g      0 t
ORIGIN

  Query Match             100.0%;  Score 8;  DB 17;  Length 28;
  Best Local Similarity   100.0%;  Pred. No. 3.2e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db     19 CGCGCGCG 26


RESULT 10
AZ627049/c
LOCUS       AZ627049                 28 bp    DNA     linear   GSS 13-DEC-2000
DEFINITION  1M0467E21R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0467E21 R, DNA sequence.
ACCESSION   AZ627049
VERSION     AZ627049.1  GI:11749239
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 28)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0467  row: E  column: 21
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 28.
FEATURES             Location/Qualifiers
     source          1. .28
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0467E21"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
```

```
BASE COUNT        0 a     14 c     14 g      0 t
ORIGIN

  Query Match             100.0%;  Score 8;  DB 17;  Length 28;
  Best Local Similarity   100.0%;  Pred. No. 3.2e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db     10 CGCGCGCG 3


RESULT 11
BI556227
LOCUS       BI556227                 30 bp    mRNA    linear   EST 05-SEP-2001
DEFINITION  603237625F1 NCI_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5290638 5',
            mRNA sequence.
ACCESSION   BI556227
VERSION     BI556227.1  GI:15443541
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 30)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11735  row: a  column: 07
            High quality sequence stop: 30.
FEATURES             Location/Qualifiers
     source          1. .30
                     /organism="Mus musculus"
                     /strain="129,C57BL/6J,FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5290638"
                     /clone_lib="NCI_CGAP_Mam3"
                     /tissue_type="tumor, gross tissue"
                     /dev_stage="10 months"
                     /lab_host="DH10B"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigators
                     providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH
                     Reference for transgenic model: Xu et al., Nature Genetics
                     22, 37-43 (1999)."
BASE COUNT        5 a      8 c     14 g      3 t
ORIGIN

  Query Match             100.0%;  Score 8;  DB 13;  Length 30;
  Best Local Similarity   100.0%;  Pred. No. 3.2e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db     22 CGCGCGCG 29


RESULT 12
BI556227/c
LOCUS       BI556227                 30 bp    mRNA    linear   EST 05-SEP-2001
DEFINITION  603237625F1 NCI_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5290638 5',
            mRNA sequence.
ACCESSION   BI556227
```

```
VERSION     BI556227.1  GI:15443541
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 30)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11735  row: a  column: 07
            High quality sequence stop: 30.
FEATURES             Location/Qualifiers
     source          1. .30
                     /organism="Mus musculus"
                     /strain="129,C57BL/6J,FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5290638"
                     /clone_lib="NCI_CGAP_Mam3"
                     /tissue_type="tumor, gross tissue"
                     /dev_stage="10 months"
                     /lab_host="DH10B"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigators
                     providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH
                     Reference for transgenic model: Xu et al., Nature Genetics
                     22, 37-43 (1999)."
BASE COUNT        5 a      8 c     14 g      3 t
ORIGIN

  Query Match             100.0%;  Score 8;  DB 13;  Length 30;
  Best Local Similarity   100.0%;  Pred. No. 3.2e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db     29 CGCGCGCG 22


RESULT 13
BF302851
LOCUS       BF302851                38 bp    mRNA    linear   EST 21-NOV-2000
DEFINITION  602032746F1 NCI_CGAP_SG2 Mus musculus cDNA clone IMAGE:4167706 5',
            mRNA sequence.
ACCESSION   BF302851
VERSION     BF302851.1  GI:11249409
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 38)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9458  row: l  column: 11
            High quality sequence stop: 38.
FEATURES             Location/Qualifiers
     source          1. .38
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4167706"
                     /clone_lib="NCI_CGAP_SG2"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: salivary gland; Vector: pCMV-SPORT6; Site_1:
                     NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo
                     dT. Average insert size 1.3 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT        6 a     10 c     21 g      1 t
ORIGIN

  Query Match             100.0%;  Score 8;  DB 12;  Length 38;
  Best Local Similarity   100.0%;  Pred. No. 3.1e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db     30 CGCGCGCG 37


RESULT 14
BF302851/c
LOCUS       BF302851                38 bp    mRNA    linear   EST 21-NOV-2000
DEFINITION  602032746F1 NCI_CGAP_SG2 Mus musculus cDNA clone IMAGE:4167706 5',
            mRNA sequence.
ACCESSION   BF302851
VERSION     BF302851.1  GI:11249409
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 38)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9458  row: l  column: 11
            High quality sequence stop: 38.
FEATURES             Location/Qualifiers
     source          1. .38
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4167706"
                     /clone_lib="NCI_CGAP_SG2"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: salivary gland; Vector: pCMV-SPORT6; Site_1:
                     NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo
                     dT. Average insert size 1.3 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT        6 a     10 c     21 g      1 t
ORIGIN

  Query Match             100.0%;  Score 8;  DB 12;  Length 38;
  Best Local Similarity   100.0%;  Pred. No. 3.1e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy      1 CGCGCGCG 8
          ||||||||
Db     37 CGCGCGCG 30


RESULT 15
BM400755
LOCUS       BM400755                  38 bp    mRNA    linear   EST 17-JAN-2002
DEFINITION  5009-0-78-H02.t.1 Chilcoat/Turkewitz cDNA (large fraction)
            Tetrahymena thermophila cDNA, mRNA sequence.
ACCESSION   BM400755
VERSION     BM400755.1  GI:18200808
KEYWORDS    EST.
SOURCE      Tetrahymena thermophila.
  ORGANISM  Tetrahymena thermophila
            Eukaryota; Alveolata; Ciliophora; Oligohymenophorea;
            Hymenostomatida; Tetrahymenina; Tetrahymena.
REFERENCE   1  (bases 1 to 38)
  AUTHORS   Turkewitz,A.P., Karrer,K.M., Jahn,C., Orias,E., Kirk,K.E., Frankel
            ,J. and Klobutcher,L.
  TITLE     EST from Tetrahymena thermophila, strain CU428.1, growing cells
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Turkewitz AP
            Molecular Genetics and Cell Biology
            University of Chicago
            920 E. 58th Street, Chicago, IL 60637, USA
            Tel: 773 702 4374
            Fax: 773 702 3172
            Email: apturkew@midway.uchicago.edu
            Seq primer: T3.
FEATURES             Location/Qualifiers
     source          1. .38
                     /organism="Tetrahymena thermophila"
                     /strain="CU428.1"
                     /db_xref="taxon:5911"
                     /clone_lib="Chilcoat/Turkewitz cDNA (large fraction)"
                     /note="Vector: BlueScript2 SK+; Details on library
                     preparation can be found in Chilcoat and Turkewitz (2001)
                     Proc. Natl. Acad. Sci USA, 98: 8709-8713."
BASE COUNT        4 a     12 c     16 g      4 t      2 others
ORIGIN

  Query Match             100.0%;  Score 8;  DB 13;  Length 38;
  Best Local Similarity   100.0%;  Pred. No. 3.1e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db      3 CGCGCGCG 10


RESULT 16
BM400755/c
LOCUS       BM400755                  38 bp    mRNA    linear   EST 17-JAN-2002
DEFINITION  5009-0-78-H02.t.1 Chilcoat/Turkewitz cDNA (large fraction)
            Tetrahymena thermophila cDNA, mRNA sequence.
ACCESSION   BM400755
VERSION     BM400755.1  GI:18200808
KEYWORDS    EST.
SOURCE      Tetrahymena thermophila.
  ORGANISM  Tetrahymena thermophila
            Eukaryota; Alveolata; Ciliophora; Oligohymenophorea;
            Hymenostomatida; Tetrahymenina; Tetrahymena.
REFERENCE   1  (bases 1 to 38)
  AUTHORS   Turkewitz,A.P., Karrer,K.M., Jahn,C., Orias,E., Kirk,K.E., Frankel
            ,J. and Klobutcher,L.
  TITLE     EST from Tetrahymena thermophila, strain CU428.1, growing cells
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Turkewitz AP
            Molecular Genetics and Cell Biology
            University of Chicago
            920 E. 58th Street, Chicago, IL 60637, USA
            Tel: 773 702 4374
            Fax: 773 702 3172
            Email: apturkew@midway.uchicago.edu
            Seq primer: T3.
FEATURES             Location/Qualifiers
     source          1. .38
                     /organism="Tetrahymena thermophila"
                     /strain="CU428.1"
                     /db_xref="taxon:5911"
                     /clone_lib="Chilcoat/Turkewitz cDNA (large fraction)"
                     /note="Vector: BlueScript2 SK+; Details on library
                     preparation can be found in Chilcoat and Turkewitz (2001)
                     Proc. Natl. Acad. Sci USA, 98: 8709-8713."
BASE COUNT        4 a     12 c     16 g      4 t      2 others
ORIGIN

  Query Match             100.0%;  Score 8;  DB 13;  Length 38;
  Best Local Similarity   100.0%;  Pred. No. 3.1e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db     10 CGCGCGCG 3


RESULT 17
BG538554
LOCUS       BG538554                  39 bp    mRNA    linear   EST 03-APR-2001
DEFINITION  602567279F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4691923 5',
            mRNA sequence.
ACCESSION   BG538554
VERSION     BG538554.1  GI:13530787
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 39)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONTECH Laboratories, Inc.
             cDNA Library Preparation: CLONTECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1511  row: n  column: 20
            High quality sequence stop: 39.
FEATURES             Location/Qualifiers
     source          1. .39
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4691923"
                     /clone_lib="NIH_MGC_77"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: lung; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); 5' and
                     3' adaptors were used in cloning as follows: 5' adaptor
                     sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C,  G, or T).  Average insert size 1.9
                     kb (range 0.5-4.0 kb). 12/15 colonies contained inserts
                     by PCR. This library was enriched for full-length clones
                     and was constructed by Clontech  Laboratories (Palo Alto,
                     CA). Note: this is a NIH_MGC Library."
BASE COUNT        6 a     13 c     17 g      3 t
ORIGIN

  Query Match             100.0%;  Score 8;  DB 12;  Length 39;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
```

```
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db     25 CGCGCGCG 32


RESULT 18
BG538554/c
LOCUS       BG538554                 39 bp    mRNA    linear   EST 03-APR-2001
DEFINITION  602567279F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4691923 5',
            mRNA sequence.
ACCESSION   BG538554
VERSION     BG538554.1  GI:13530787
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 39)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONTECH Laboratories, Inc.
             cDNA Library Preparation: CLONTECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1511  row: n  column: 20
            High quality sequence stop: 39.
FEATURES             Location/Qualifiers
     source          1. .39
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4691923"
                     /clone_lib="NIH_MGC_77"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: lung; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); 5' and
                     3' adaptors were used in cloning as follows: 5' adaptor
                     sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C,  G, or T).  Average insert size 1.9
                     kb (range 0.5-4.0 kb). 12/15 colonies contained inserts
                     by PCR. This library was enriched for full-length clones
                     and was constructed by Clontech  Laboratories (Palo Alto,
                     CA). Note: this is a NIH_MGC Library."
BASE COUNT        6 a     13 c     17 g      3 t
ORIGIN

  Query Match             100.0%;  Score 8;  DB 12;  Length 39;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db     32 CGCGCGCG 25


RESULT 19
AZ431922
LOCUS       AZ431922                 40 bp    DNA     linear   GSS 03-OCT-2000
DEFINITION  1M0217D21F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0217D21 F, DNA sequence.
ACCESSION   AZ431922
VERSION     AZ431922.1  GI:10556031
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 40)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0217  row: D  column: 21
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 40.
FEATURES             Location/Qualifiers
     source          1. .40
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0217D21"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT        1 a     20 c     14 g      5 t
ORIGIN

  Query Match             100.0%;  Score 8;  DB 17;  Length 40;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db     24 CGCGCGCG 31


RESULT 20
AZ431922/c
LOCUS       AZ431922                 40 bp    DNA     linear   GSS 03-OCT-2000
DEFINITION  1M0217D21F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0217D21 F, DNA sequence.
ACCESSION   AZ431922
VERSION     AZ431922.1  GI:10556031
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
```

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 40)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0217  row: D  column: 21
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 40.
FEATURES             Location/Qualifiers
     source          1. .40
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0217D21"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT        1 a     20 c     14 g      5 t
ORIGIN

  Query Match          .    100.0%;  Score 8;  DB 17;  Length 40;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db     31 CGCGCGCG 24


RESULT 21
BH796426
LOCUS       BH796426                 40 bp    DNA     linear   GSS 25-APR-2002
DEFINITION  1008094B12.1EL_x1 1008 - RescueMu Grid I Zea mays genomic, DNA
            sequence.
ACCESSION   BH796426
VERSION     BH796426.1  GI:20305035
KEYWORDS    GSS.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 40)
  AUTHORS   Walbot,V.
  TITLE     Maize genomic sequences found using engineered RescueMu transposon
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Very probable ligation site of ends cut by single endonuclease.
            Reverse complemented post-ligation sequence from source sequence.
            Plate: 1008094  row: 26
            Class: transposon-tagged.
FEATURES             Location/Qualifiers
     source          1. .40
                     /organism="Zea mays"
                     /cultivar="mixed background W23/A188/B73"
                     /db_xref="taxon:4577"
                     /clone_lib="1008 - RescueMu Grid I"
                     /tissue_type="leaf"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: leaf; Vector: RescueMu (engineered from
                     pBlueScript backbone); Site_1: BamHI; Site_2: BglII;
                     RescueMu is a 4.9 kb, modified maize Mu transposon
                     designed to allow plasmid rescue from total genomic DNA.
                     Mu elements insert preferentially into transcription
                     units. For more information on RescueMu, go to the  web
                     site www.zmdb.iastate.edu and follow the links for
                     'RescueMu.' Grid I was grown at Berkeley in 2001. DNA was
                     extracted from leaf punches,  double digested using BamHI
                     and BglII, and ligated to form circular plasmids.  DH10B
                     cells were transformed and then screened on LB plates with
                     ampicillin."
BASE COUNT        2 a     15 c     13 g     10 t
ORIGIN

  Query Match          .    100.0%;  Score 8;  DB 17;  Length 40;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db     28 CGCGCGCG 35


RESULT 22
BH796426/c
LOCUS       BH796426                 40 bp    DNA     linear   GSS 25-APR-2002
DEFINITION  1008094B12.1EL_x1 1008 - RescueMu Grid I Zea mays genomic, DNA
            sequence.
ACCESSION   BH796426
VERSION     BH796426.1  GI:20305035
KEYWORDS    GSS.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 40)
  AUTHORS   Walbot,V.
  TITLE     Maize genomic sequences found using engineered RescueMu transposon
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
```

```
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Very probable ligation site of ends cut by single endonuclease.
            Reverse complemented post-ligation sequence from source sequence.
            Plate: 1008094  row: 26
            Class: transposon-tagged.
FEATURES                Location/Qualifiers
     source             1. .40
                        /organism="Zea mays"
                        /cultivar="mixed background W23/A188/B73"
                        /db_xref="taxon:4577"
                        /clone_lib="1008 - RescueMu Grid I"
                        /tissue_type="leaf"
                        /dev_stage="adult"
                        /lab_host="DH10B"
                        /note="Organ: leaf; Vector: RescueMu (engineered from
                        pBlueScript backbone); Site_1: BamHI; Site_2: BglII;
                        RescueMu is a 4.9 kb, modified maize Mu transposon
                        designed to allow plasmid rescue from total genomic DNA.
                        Mu elements insert preferentially into transcription
                        units. For more information on RescueMu, go to the  web
                        site www.zmdb.iastate.edu and follow the links for
                        'RescueMu.' Grid I was grown at Berkeley in 2001. DNA was
                        extracted from leaf punches,  double digested using BamHI
                        and BglII, and ligated to form circular plasmids.  DH10B
                        cells were transformed and then screened on LB plates with
                        ampicillin."
BASE COUNT        2 a     15 c     13 g     10 t
ORIGIN

  Query Match             100.0%;  Score 8;  DB 17;  Length 40;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db     35 CGCGCGCG 28


RESULT 23
BH639962
LOCUS       BH639962                 41 bp    DNA     linear   GSS 14-FEB-2002
DEFINITION  1008033A02.2EL_y1 1008 - RescueMu Grid I Zea mays genomic, DNA
            sequence.
ACCESSION   BH639962
VERSION     BH639962.1  GI:18665781
KEYWORDS    GSS.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 41)
  AUTHORS   Walbot,V.
  TITLE     Maize genomic sequences found using engineered RescueMu transposon
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Possible ligation site of ends cut by 2 different endonucleases.
            Reverse complemented post-ligation sequence from source sequence.
            Plate: 1008033  row: 34
            Class: transposon-tagged.
FEATURES                Location/Qualifiers
     source             1. .41
                        /organism="Zea mays"
                        /cultivar="mixed background W23/A188/B73"
                        /db_xref="taxon:4577"
                        /clone_lib="1008 - RescueMu Grid I"
                        /tissue_type="leaf"
                        /dev_stage="adult"
                        /lab_host="DH10B"
                        /note="Organ: leaf; Vector: RescueMu (engineered from
                        pBlueScript backbone); Site_1: BamHI; Site_2: BglII;
                        RescueMu is a 4.9 kb, modified maize Mu transposon
                        designed to allow plasmid rescue from total genomic DNA.
                        Mu elements insert preferentially into transcription
                        units. For more information on RescueMu, go to the  web
                        site www.zmdb.iastate.edu and follow the links for
                        'RescueMu.' Grid I was grown at Berkeley in 2001. DNA was
                        extracted from leaf punches,  double digested using BamHI
                        and BglII, and ligated to form circular plasmids.  DH10B
                        cells were transformed and then screened on LB plates with
                        ampicillin."
BASE COUNT        6 a     12 c     21 g      2 t
ORIGIN

  Query Match             100.0%;  Score 8;  DB 17;  Length 41;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db     15 CGCGCGCG 22


RESULT 24
BH639962/c
LOCUS       BH639962                 41 bp    DNA     linear   GSS 14-FEB-2002
DEFINITION  1008033A02.2EL_y1 1008 - RescueMu Grid I Zea mays genomic, DNA
            sequence.
ACCESSION   BH639962
VERSION     BH639962.1  GI:18665781
KEYWORDS    GSS.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 41)
  AUTHORS   Walbot,V.
  TITLE     Maize genomic sequences found using engineered RescueMu transposon
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Possible ligation site of ends cut by 2 different endonucleases.
            Reverse complemented post-ligation sequence from source sequence.
            Plate: 1008033  row: 34
            Class: transposon-tagged.
FEATURES                Location/Qualifiers
     source             1. .41
                        /organism="Zea mays"
                        /cultivar="mixed background W23/A188/B73"
                        /db_xref="taxon:4577"
                        /clone_lib="1008 - RescueMu Grid I"
                        /tissue_type="leaf"
                        /dev_stage="adult"
                        /lab_host="DH10B"
                        /note="Organ: leaf; Vector: RescueMu (engineered from
                        pBlueScript backbone); Site_1: BamHI; Site_2: BglII;
                        RescueMu is a 4.9 kb, modified maize Mu transposon
                        designed to allow plasmid rescue from total genomic DNA.
                        Mu elements insert preferentially into transcription
                        units. For more information on RescueMu, go to the  web
                        site www.zmdb.iastate.edu and follow the links for
                        'RescueMu.' Grid I was grown at Berkeley in 2001. DNA was
                        extracted from leaf punches,  double digested using BamHI
```

```
                    and BglII, and ligated to form circular plasmids.  DH10B
                    cells were transformed and then screened on LB plates with
                    ampicillin."
BASE COUNT        6 a     12 c     21 g      2 t
ORIGIN

  Query Match             100.0%;  Score 8;  DB 17;  Length 41;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db     22 CGCGCGCG 15


RESULT 25
BG915507
LOCUS       BG915507                 43 bp    mRNA    linear   EST 05-JUN-2001
DEFINITION  602815734F1 NCI_CGAP_Mam4 Mus musculus cDNA clone IMAGE:4945113 5',
            mRNA sequence.
ACCESSION   BG915507
VERSION     BG915507.1  GI:14295983
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 43)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furth
            Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10891  row: d  column: 10
            High quality sequence stop: 40.
FEATURES             Location/Qualifiers
     source          1. .43
                     /organism="Mus musculus"
                     /strain="NMRI"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4945113"
                     /clone_lib="NCI_CGAP_Mam4"
                     /tissue_type="tumor, gross tissue"
                     /dev_stage="5 months"
                     /lab_host="DH10B"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigators
                     providing samples: Lothar Hennighausen/Priscilla Furth,
                     NIH Reference for transgenic model: Li et al., Cell Growth
                     and  Differentiation 7, 3-11 (1996)."
BASE COUNT        4 a     16 c     19 g      4 t
ORIGIN

  Query Match             100.0%;  Score 8;  DB 13;  Length 43;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db     30 CGCGCGCG 37


RESULT 26
BG915507/c
LOCUS       BG915507                 43 bp    mRNA    linear   EST 05-JUN-2001
DEFINITION  602815734F1 NCI_CGAP_Mam4 Mus musculus cDNA clone IMAGE:4945113 5',
            mRNA sequence.
ACCESSION   BG915507
VERSION     BG915507.1  GI:14295983
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 43)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furth
            Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10891  row: d  column: 10
            High quality sequence stop: 40.
FEATURES             Location/Qualifiers
     source          1. .43
                     /organism="Mus musculus"
                     /strain="NMRI"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4945113"
                     /clone_lib="NCI_CGAP_Mam4"
                     /tissue_type="tumor, gross tissue"
                     /dev_stage="5 months"
                     /lab_host="DH10B"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigators
                     providing samples: Lothar Hennighausen/Priscilla Furth,
                     NIH Reference for transgenic model: Li et al., Cell Growth
                     and  Differentiation 7, 3-11 (1996)."
BASE COUNT        4 a     16 c     19 g      4 t
ORIGIN

  Query Match             100.0%;  Score 8;  DB 13;  Length 43;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db     37 CGCGCGCG 30


RESULT 27
BH629051
LOCUS       BH629051                 44 bp    DNA     linear   GSS 30-JAN-2002
DEFINITION  1007076B07.2EL_y1 1007 - RescueMu Grid H Zea mays genomic, DNA
            sequence.
ACCESSION   BH629051
VERSION     BH629051.1  GI:18442302
KEYWORDS    GSS.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 44)
  AUTHORS   Walbot,V.
  TITLE     Maize genomic sequences found using engineered RescueMu transposon
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
```

```
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Possible ligation site of ends cut by 2 different endonucleases.
            Reverse complemented post-ligation sequence from source sequence.
            Plate: 1007076  column: 36
            Class: transposon-tagged.
FEATURES                Location/Qualifiers
     source             1. .44
                        /organism="Zea mays"
                        /cultivar="mixed background W23/A188/B73"
                        /db_xref="taxon:4577"
                        /clone_lib="1007 - RescueMu Grid H"
                        /tissue_type="leaf"
                        /dev_stage="adult"
                        /lab_host="DH10B"
                        /note="Organ: leaf; Vector: RescueMu (engineered from
                        pBlueScript backbone); Site_1: BamHI; Site_2: BglII;
                        RescueMu is a 4.9 kb, modified maize Mu transposon
                        designed to allow plasmid rescue from total genomic DNA.
                        Mu elements insert preferentially into transcription
                        units.  For more information on RescueMu, go to the  web
                        site 'www.zmdb.iastate.edu' and follow the links for
                        'RescueMu.'  Grid H was grown at Berkeley in 2001.  DNA
                        was extracted from leaf punches,  double digested using
                        BamHI and BglII, and ligated to form circular plasmids.
                        DH10B cells were transformed and then screened on LB
                        plates with ampicillin."
BASE COUNT       15 a     20 c      6 g      3 t
ORIGIN

  Query Match             100.0%;  Score 8;  DB 17;  Length 44;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db      8 CGCGCGCG 15


RESULT 28
BH629051/c
LOCUS       BH629051                 44 bp    DNA     linear   GSS 30-JAN-2002
DEFINITION  1007076B07.2EL_y1 1007 - RescueMu Grid H Zea mays genomic, DNA
            sequence.
ACCESSION   BH629051
VERSION     BH629051.1  GI:18442302
KEYWORDS    GSS.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 44)
  AUTHORS   Walbot,V.
  TITLE     Maize genomic sequences found using engineered RescueMu transposon
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Possible ligation site of ends cut by 2 different endonucleases.
            Reverse complemented post-ligation sequence from source sequence.
            Plate: 1007076  column: 36
            Class: transposon-tagged.
FEATURES                Location/Qualifiers
     source             1. .44
                        /organism="Zea mays"
                        /cultivar="mixed background W23/A188/B73"
                        /db_xref="taxon:4577"
                        /clone_lib="1007 - RescueMu Grid H"
                        /tissue_type="leaf"
                        /dev_stage="adult"
                        /lab_host="DH10B"
                        /note="Organ: leaf; Vector: RescueMu (engineered from
                        pBlueScript backbone); Site_1: BamHI; Site_2: BglII;
                        RescueMu is a 4.9 kb, modified maize Mu transposon
                        designed to allow plasmid rescue from total genomic DNA.
                        Mu elements insert preferentially into transcription
                        units.  For more information on RescueMu, go to the  web
                        site 'www.zmdb.iastate.edu' and follow the links for
                        'RescueMu.'  Grid H was grown at Berkeley in 2001.  DNA
                        was extracted from leaf punches,  double digested using
                        BamHI and BglII, and ligated to form circular plasmids.
                        DH10B cells were transformed and then screened on LB
                        plates with ampicillin."
BASE COUNT       15 a     20 c      6 g      3 t
ORIGIN

  Query Match             100.0%;  Score 8;  DB 17;  Length 44;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db      15 CGCGCGCG 8


RESULT 29
AZ514564
LOCUS       AZ514564                 45 bp    DNA     linear   GSS 05-OCT-2000
DEFINITION  1M0361N10F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0361N10 F, DNA sequence.
ACCESSION   AZ514564
VERSION     AZ514564.1  GI:10695796
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 45)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0361  row: N  column: 10
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 45.
FEATURES                Location/Qualifiers
     source             1. .45
                        /organism="Mus musculus"
                        /strain="C57BL/6J"
                        /db_xref="taxon:10090"
                        /clone="UUGC1M0361N10"
                        /clone_lib="Mouse 10kb plasmid UUGC1M library"
                        /sex="Male"
                        /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                        /note="Vector: PWD42nv; Purified genomic DNA from M.
```

musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

```
BASE COUNT       17 a     22 c      6 g      0 t
ORIGIN

  Query Match             100.0%;  Score 8;  DB 17;  Length 45;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db      36 CGCGCGCG 43


RESULT 30
AZ514564/c
LOCUS       AZ514564                 45 bp    DNA     linear   GSS 05-OCT-2000
DEFINITION  1M0361N10F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0361N10 F, DNA sequence.
ACCESSION   AZ514564
VERSION     AZ514564.1  GI:10695796
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 45)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0361  row: N  column: 10
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 45.
FEATURES             Location/Qualifiers
     source          1. .45
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0361N10"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
```

musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

```
BASE COUNT       17 a     22 c      6 g      0 t
ORIGIN

  Query Match             100.0%;  Score 8;  DB 17;  Length 45;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db      45 CGCGCGCG 38


RESULT 31
AZ400633
LOCUS       AZ400633                 47 bp    DNA     linear   GSS 03-OCT-2000
DEFINITION  1M0167F06F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0167F06 F, DNA sequence.
ACCESSION   AZ400633
VERSION     AZ400633.1  GI:10515707
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 47)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0167  row: F  column: 06
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 47.
FEATURES             Location/Qualifiers
     source          1. .47
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0167F06"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
```

musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 2 a 15 c 23 g 7 t
ORIGIN

Query Match 100.0%; Score 8; DB 17; Length 47;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 CGCGCGCG 8
           ||||||||
Db      15 CGCGCGCG 22
```

RESULT 32
AZ400633/c
LOCUS AZ400633 47 bp DNA linear GSS 03-OCT-2000
DEFINITION 1M0167F06F Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0167F06 F, DNA sequence.
ACCESSION AZ400633
VERSION AZ400633.1 GI:10515707
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 47)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0167 row: F column: 06
Seq primer: CGTTGTAAAACGACGGCCAGT
Class: plasmid ends
High quality sequence stop: 47.
FEATURES Location/Qualifiers
source 1. .47
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0167F06"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M.

musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 2 a 15 c 23 g 7 t
ORIGIN

Query Match 100.0%; Score 8; DB 17; Length 47;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 CGCGCGCG 8
           ||||||||
Db      24 CGCGCGCG 17
```

RESULT 33
AZ585161
LOCUS AZ585161 47 bp DNA linear GSS 13-DEC-2000
DEFINITION 1M0390G12F Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0390G12 F, DNA sequence.
ACCESSION AZ585161
VERSION AZ585161.1 GI:11706772
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 47)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0390 row: G column: 12
Seq primer: CGTTGTAAAACGACGGCCAGT
Class: plasmid ends
High quality sequence stop: 47.
FEATURES Location/Qualifiers
source 1. .47
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0390G12"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M.

musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 14 a 22 c 11 g 0 t
ORIGIN

Query Match 100.0%; Score 8; DB 17; Length 47;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 CGCGCGCG 8
          ||||||||
Db      2 CGCGCGCG 9
```

RESULT 34
AZ585161/c
LOCUS AZ585161 47 bp DNA linear GSS 13-DEC-2000
DEFINITION 1M0390G12F Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0390G12 F, DNA sequence.
ACCESSION AZ585161
VERSION AZ585161.1 GI:11706772
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 47)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0390 row: G column: 12
Seq primer: CGTTGTAAAACGACGGCCAGT
Class: plasmid ends
High quality sequence stop: 47.

```
FEATURES             Location/Qualifiers
     source          1. .47
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0390G12"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
```

musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 14 a 22 c 11 g 0 t
ORIGIN

Query Match 100.0%; Score 8; DB 17; Length 47;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 CGCGCGCG 8
          ||||||||
Db      9 CGCGCGCG 2
```

RESULT 35
AZ764402
LOCUS AZ764402 48 bp DNA linear GSS 16-FEB-2001
DEFINITION 1M0560O18F Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0560O18 F, DNA sequence.
ACCESSION AZ764402
VERSION AZ764402.1 GI:12879326
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 48)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0560 row: O column: 18
Seq primer: CGTTGTAAAACGACGGCCAGT
Class: plasmid ends
High quality sequence stop: 48.

```
FEATURES             Location/Qualifiers
     source          1. .48
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0560O18"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
```

musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

```
BASE COUNT        1 a      10 c      24 g      13 t
ORIGIN

  Query Match             100.0%;  Score 8;  DB 17;  Length 48;
  Best Local Similarity   100.0%;  Pred. No. 2.9e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       2 CGCGCGCG 9
```

```
RESULT 36
AZ764402/c
LOCUS       AZ764402                48 bp    DNA     linear   GSS 16-FEB-2001
DEFINITION  1M0560018F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0560018 F, DNA sequence.
ACCESSION   AZ764402
VERSION     AZ764402.1  GI:12879326
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 48)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0560  row: O  column: 18
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 48.
FEATURES             Location/Qualifiers
     source          1. .48
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0560018"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
```

musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

```
BASE COUNT        1 a      10 c      24 g      13 t
ORIGIN

  Query Match             100.0%;  Score 8;  DB 17;  Length 48;
  Best Local Similarity   100.0%;  Pred. No. 2.9e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       9 CGCGCGCG 2
```

```
RESULT 37
AA948394
LOCUS       AA948394                49 bp    mRNA    linear   EST 23-JUN-1998
DEFINITION  on52b09.s1 NCI_CGAP_Co8 Homo sapiens cDNA clone IMAGE:1560281 3'
            similar to SW:CA11_CHICK P02457 PROCOLLAGEN ALPHA 1(I) CHAIN
            PRECURSOR. ;, mRNA sequence.
ACCESSION   AA948394
VERSION     AA948394.1  GI:3109647
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 49)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Trace considered overall poor quality
            Insert Length: 1085   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1. .49
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1560281"
                     /clone_lib="NCI_CGAP_Co8"
                     /tissue_type="adenocarcinoma"
                     /lab_host="DH10B"
                     /note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; 1st strand cDNA was prepared from
```

```
                    colon adenocarcinoma, and was then primed with a Not I -
                    oligo(dT) primer. Double-stranded cDNA was ligated to Eco
                    RI adaptors (Pharmacia), digested with Not I and cloned
                    into the Not I and Eco RI sites of the modified pT7T3
                    vector.  Library is normalized. Library was constructed by
                    Bento Soares and M. Fatima Bonaldo. "
BASE COUNT        5 a     22 c     18 g      4 t
ORIGIN

  Query Match             100.0%;  Score 8;  DB 9;  Length 49;
  Best Local Similarity   100.0%;  Pred. No. 2.9e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db     31 CGCGCGCG 38


RESULT 38
AA948394/c
LOCUS       AA948394                 49 bp    mRNA    linear   EST 23-JUN-1998
DEFINITION  on52b09.s1 NCI_CGAP_Co8 Homo sapiens cDNA clone IMAGE:1560281 3'
            similar to SW:CA11_CHICK P02457 PROCOLLAGEN ALPHA 1(I) CHAIN
            PRECURSOR. ;, mRNA sequence.
ACCESSION   AA948394
VERSION     AA948394.1  GI:3109647
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 49)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Trace considered overall poor quality
            Insert Length: 1085    Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1. .49
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1560281"
                     /clone_lib="NCI_CGAP_Co8"
                     /tissue_type="adenocarcinoma"
                     /lab_host="DH10B"
                     /note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; 1st strand cDNA was prepared from
                     colon adenocarcinoma, and was then primed with a Not I -
                     oligo(dT) primer. Double-stranded cDNA was ligated to Eco
                     RI adaptors (Pharmacia), digested with Not I and cloned
                     into the Not I and Eco RI sites of the modified pT7T3
                     vector.  Library is normalized. Library was constructed by
                     Bento Soares and M. Fatima Bonaldo. "
BASE COUNT        5 a     22 c     18 g      4 t
ORIGIN

  Query Match             100.0%;  Score 8;  DB 9;  Length 49;
  Best Local Similarity   100.0%;  Pred. No. 2.9e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db     38 CGCGCGCG 31


RESULT 39
AI440059
LOCUS       AI440059                 49 bp    mRNA    linear   EST 09-MAR-1999
DEFINITION  ti61f10.x1 NCI_CGAP_Lym12 Homo sapiens cDNA clone IMAGE:2134987 3'
            similar to SW:PRPL_HUMAN P10162 SALIVARY PROLINE-RICH PROTEIN PO
            ;contains element TAR1 repetitive element ;, mRNA sequence.
ACCESSION   AI440059
VERSION     AI440059.1  GI:4308506
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 49)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Life Technologies catalog #: 11547-015
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Trace considered overall poor quality
            Seq primer: -40UP from Gibco
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1. .49
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2134987"
                     /clone_lib="NCI_CGAP_Lym12"
                     /tissue_type="lymphoma, follicular mixed small and large
                     cell"
                     /lab_host="DH10B"
                     /note="Organ: lymph node; Vector: pCMV-SPORT6; Site_1:
                     SalI; Site_2: NotI; Cloned unidirectionally.  Primer:
                     Oligo dT.  Average insert size 1.25 kb. Life Technologies
                     catalog #: 11547-015"
BASE COUNT       13 a     20 c     14 g      2 t
ORIGIN

  Query Match             100.0%;  Score 8;  DB 9;  Length 49;
  Best Local Similarity   100.0%;  Pred. No. 2.9e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db     24 CGCGCGCG 31


RESULT 40
AI440059/c
LOCUS       AI440059                 49 bp    mRNA    linear   EST 09-MAR-1999
DEFINITION  ti61f10.x1 NCI_CGAP_Lym12 Homo sapiens cDNA clone IMAGE:2134987 3'
            similar to SW:PRPL_HUMAN P10162 SALIVARY PROLINE-RICH PROTEIN PO
            ;contains element TAR1 repetitive element ;, mRNA sequence.
ACCESSION   AI440059
VERSION     AI440059.1  GI:4308506
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
REFERENCE   1  (bases 1 to 49)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Life Technologies catalog #: 11547-015
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Trace considered overall poor quality
            Seq primer: -40UP from Gibco
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1. .49
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2134987"
                     /clone_lib="NCI_CGAP_Lym12"
                     /tissue_type="lymphoma, follicular mixed small and large
                     cell"
                     /lab_host="DH10B"
                     /note="Organ: lymph node; Vector: pCMV-SPORT6; Site_1:
                     SalI; Site_2: NotI; Cloned unidirectionally.  Primer:
                     Oligo dT.  Average insert size 1.25 kb. Life Technologies
                     catalog #: 11547-015"
BASE COUNT       13 a     20 c     14 g      2 t
ORIGIN

  Query Match             100.0%;  Score 8;  DB 9;  Length 49;
  Best Local Similarity   100.0%;  Pred. No. 2.9e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db     31 CGCGCGCG 24


RESULT 41
AI565007
LOCUS       AI565007                  49 bp    mRNA    linear   EST 13-MAY-1999
DEFINITION  tq53d09.x1 NCI_CGAP_Ut1 Homo sapiens cDNA clone IMAGE:2212529 3'
            similar to TR:Q04154 Q04154 SALIVARY PROLINE-RICH PROTEIN RP15
            PRECURSOR. ;contains PTR5.t3 MSR1 repetitive element ;, mRNA
            sequence.
ACCESSION   AI565007
VERSION     AI565007.1  GI:4523464
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 49)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Trace considered overall poor quality
            Insert Length: 366    Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 1
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .49
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2212529"
                     /clone_lib="NCI_CGAP_Ut1"
                     /tissue_type="well-differentiated endometrial
                     adenocarcinoma, 7 pooled tumors"
                     /lab_host="DH10B"
                     /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.75 kb. Life Technologies catalog #:
                     11538-014"
BASE COUNT       10 a     21 c     18 g      0 t
ORIGIN

  Query Match             100.0%;  Score 8;  DB 9;  Length 49;
  Best Local Similarity   100.0%;  Pred. No. 2.9e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db      2 CGCGCGCG 9


RESULT 42
AI565007/c
LOCUS       AI565007                  49 bp    mRNA    linear   EST 13-MAY-1999
DEFINITION  tq53d09.x1 NCI_CGAP_Ut1 Homo sapiens cDNA clone IMAGE:2212529 3'
            similar to TR:Q04154 Q04154 SALIVARY PROLINE-RICH PROTEIN RP15
            PRECURSOR. ;contains PTR5.t3 MSR1 repetitive element ;, mRNA
            sequence.
ACCESSION   AI565007
VERSION     AI565007.1  GI:4523464
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 49)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Trace considered overall poor quality
            Insert Length: 366    Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 1
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .49
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2212529"
                     /clone_lib="NCI_CGAP_Ut1"
                     /tissue_type="well-differentiated endometrial
                     adenocarcinoma, 7 pooled tumors"
                     /lab_host="DH10B"
                     /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI;
```

Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.75 kb. Life Technologies catalog #: 11538-014"
BASE COUNT 10 a 21 c 18 g 0 t
ORIGIN

Query Match 100.0%; Score 8; DB 9; Length 49;
Best Local Similarity 100.0%; Pred. No. 2.9e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGCGCGCG 8
||||||||
Db 9 CGCGCGCG 2

RESULT 43
AZ456464
LOCUS AZ456464 49 bp DNA linear GSS 04-OCT-2000
DEFINITION 1M0259L12F Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0259L12 F, DNA sequence.
ACCESSION AZ456464
VERSION AZ456464.1 GI:10614505
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 49)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0259 row: L column: 12
Seq primer: CGTTGTAAAACGACGGCCAGT
Class: plasmid ends
High quality sequence stop: 49.
FEATURES Location/Qualifiers
source 1. .49
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0259L12"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT 15 a 24 c 9 g 1 t
ORIGIN

Query Match 100.0%; Score 8; DB 17; Length 49;
Best Local Similarity 100.0%; Pred. No. 2.9e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGCGCGCG 8
||||||||
Db 41 CGCGCGCG 48

RESULT 44
AZ456464/c
LOCUS AZ456464 49 bp DNA linear GSS 04-OCT-2000
DEFINITION 1M0259L12F Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0259L12 F, DNA sequence.
ACCESSION AZ456464
VERSION AZ456464.1 GI:10614505
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 49)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0259 row: L column: 12
Seq primer: CGTTGTAAAACGACGGCCAGT
Class: plasmid ends
High quality sequence stop: 49.
FEATURES Location/Qualifiers
source 1. .49
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0259L12"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to

adaptored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT 15 a 24 c 9 g 1 t
ORIGIN

Query Match 100.0%; Score 8; DB 17; Length 49;
Best Local Similarity 100.0%; Pred. No. 2.9e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGCGCGCG 8
||||||||
Db 46 CGCGCGCG 39

RESULT 45
AZ773338
LOCUS AZ773338 49 bp DNA linear GSS 16-FEB-2001
DEFINITION 1M0584L13R Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0584L13 R, DNA sequence.
ACCESSION AZ773338
VERSION AZ773338.1 GI:12897597
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 49)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0584 row: L column: 13
Seq primer: CACACAGGAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 49.
FEATURES Location/Qualifiers
source 1. .49
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0584L13"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to

adaptored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT 2 a 8 c 24 g 15 t
ORIGIN

Query Match 100.0%; Score 8; DB 17; Length 49;
Best Local Similarity 100.0%; Pred. No. 2.9e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGCGCGCG 8
||||||||
Db 24 CGCGCGCG 31

Search completed: March 14, 2003, 04:32:27
Job time : 1341 secs

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: March 14, 2003, 03:39:03 ; Search time 60 Seconds
(without alignments)
93.582 Million cell updates/sec

Title: CGCGCGCG
Perfect score: 8
Sequence: 1 cgcgcgcg 8

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 501302 seqs, 350932545 residues

Total number of hits satisfying chosen parameters: 1002604

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	8	100.0	19	9	US-09-888-326-342	Sequence 342, App
c	2	8	100.0	19	9	US-09-888-326-342	Sequence 342, App
	3	8	100.0	20	9	US-09-888-326-192	Sequence 192, App
c	4	8	100.0	20	9	US-09-888-326-192	Sequence 192, App
	5	8	100.0	20	9	US-09-888-326-193	Sequence 193, App
c	6	8	100.0	20	9	US-09-888-326-193	Sequence 193, App
	7	8	100.0	20	9	US-09-888-326-607	Sequence 607, App
c	8	8	100.0	20	9	US-09-888-326-607	Sequence 607, App
	9	8	100.0	21	10	US-09-885-441-54	Sequence 54, Appl
c	10	8	100.0	21	10	US-09-885-441-54	Sequence 54, Appl
	11	8	100.0	25	10	US-09-728-721-68	Sequence 68, Appl
c	12	8	100.0	25	10	US-09-728-721-68	Sequence 68, Appl
	13	8	100.0	25	10	US-09-841-879B-16	Sequence 16, Appl
c	14	8	100.0	25	10	US-09-841-879B-16	Sequence 16, Appl
	15	8	100.0	30	10	US-09-682-597-4	Sequence 4, Appli
c	16	8	100.0	30	10	US-09-682-597-4	Sequence 4, Appli
	17	8	100.0	36	10	US-09-828-034-5	Sequence 5, Appli
c	18	8	100.0	36	10	US-09-828-034-5	Sequence 5, Appli
	19	8	100.0	39	9	US-09-733-042-31	Sequence 31, Appl
c	20	8	100.0	39	9	US-09-733-042-31	Sequence 31, Appl
	21	8	100.0	45	9	US-10-007-132-45	Sequence 45, Appl
c	22	8	100.0	45	9	US-10-007-132-45	Sequence 45, Appl
	23	8	100.0	55	10	US-09-790-417-19	Sequence 19, Appl
c	24	8	100.0	55	10	US-09-790-417-19	Sequence 19, Appl
	25	8	100.0	60	10	US-09-899-381-37	Sequence 37, Appl
c	26	8	100.0	60	10	US-09-899-381-37	Sequence 37, Appl
	27	8	100.0	61	12	US-10-014-973A-19	Sequence 19, Appl
c	28	8	100.0	61	12	US-10-014-973A-19	Sequence 19, Appl
	29	8	100.0	64	9	US-10-057-940-8	Sequence 8, Appli
c	30	8	100.0	64	9	US-10-057-940-8	Sequence 8, Appli
	31	8	100.0	68	10	US-09-473-872-54	Sequence 54, Appl
c	32	8	100.0	68	10	US-09-473-872-54	Sequence 54, Appl
	33	8	100.0	72	9	US-09-835-976B-122	Sequence 122, App
c	34	8	100.0	72	9	US-09-835-976B-122	Sequence 122, App
	35	8	100.0	74	9	US-09-835-976B-119	Sequence 119, App
c	36	8	100.0	74	9	US-09-835-976B-119	Sequence 119, App
	37	8	100.0	76	9	US-09-835-976B-127	Sequence 127, App
c	38	8	100.0	76	9	US-09-835-976B-127	Sequence 127, App
	39	8	100.0	76	9	US-09-835-976B-129	Sequence 129, App
c	40	8	100.0	76	9	US-09-835-976B-129	Sequence 129, App
	41	8	100.0	78	9	US-09-835-976B-126	Sequence 126, App
c	42	8	100.0	78	9	US-09-835-976B-126	Sequence 126, App
	43	8	100.0	78	9	US-09-835-976B-128	Sequence 128, App
c	44	8	100.0	78	9	US-09-835-976B-128	Sequence 128, App
	45	8	100.0	80	9	US-09-835-976B-124	Sequence 124, App

ALIGNMENTS

RESULT 1
US-09-888-326-342
; Sequence 342, Application US/09888326
; Publication No. US20030026801A1
; GENERAL INFORMATION:
; APPLICANT: Weiner, George
; APPLICANT: Hartmann, Gunther
; TITLE OF INVENTION: Methods for Enhancing Antibody-Induced
; TITLE OF INVENTION: Cell Lysis and Treating Cancer
; FILE REFERENCE: C1039/7052 (AWS)
; CURRENT APPLICATION NUMBER: US/09/888,326
; CURRENT FILING DATE: 2001-06-22
; PRIOR APPLICATION NUMBER: US 60/213,346
; PRIOR FILING DATE: 2000-06-22
; NUMBER OF SEQ ID NOS: 848
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 342
; LENGTH: 19
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic oligonucleotide
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: phosphodiester backbone
US-09-888-326-342

Query Match 100.0%; Score 8; DB 9; Length 19;
Best Local Similarity 100.0%; Pred. No. 1.5e+04;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 CGCGCGCG 8
             ||||||||
Db      10 CGCGCGCG 17
```

RESULT 2
US-09-888-326-342/c
; Sequence 342, Application US/09888326
; Publication No. US20030026801A1
; GENERAL INFORMATION:
; APPLICANT: Weiner, George

```
;  APPLICANT:  Hartmann, Gunther
;  TITLE OF INVENTION: Methods for Enhancing Antibody-Induced
;  TITLE OF INVENTION:  Cell Lysis and Treating Cancer
;  FILE REFERENCE: C1039/7052 (AWS)
;  CURRENT APPLICATION NUMBER: US/09/888,326
;  CURRENT FILING DATE:  2001-06-22
;  PRIOR APPLICATION NUMBER: US 60/213,346
;  PRIOR FILING DATE: 2000-06-22
;  NUMBER OF SEQ ID NOS: 848
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 342
;   LENGTH: 19
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic oligonucleotide
;   NAME/KEY: misc_feature
;   LOCATION: (0)...(0)
;   OTHER INFORMATION: phosphodiester backbone
US-09-888-326-342

  Query Match             100.0%;  Score 8;  DB 9;  Length 19;
  Best Local Similarity   100.0%;  Pred. No. 1.5e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
            ||||||||
Db       9 CGCGCGCG 2


RESULT 3
US-09-888-326-192
; Sequence 192, Application US/09888326
; Publication No. US20030026801A1
; GENERAL INFORMATION:
;  APPLICANT: Weiner, George
;  APPLICANT:  Hartmann, Gunther
;  TITLE OF INVENTION: Methods for Enhancing Antibody-Induced
;  TITLE OF INVENTION:  Cell Lysis and Treating Cancer
;  FILE REFERENCE: C1039/7052 (AWS)
;  CURRENT APPLICATION NUMBER: US/09/888,326
;  CURRENT FILING DATE:  2001-06-22
;  PRIOR APPLICATION NUMBER: US 60/213,346
;  PRIOR FILING DATE: 2000-06-22
;  NUMBER OF SEQ ID NOS: 848
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 192
;   LENGTH: 20
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic oligonucleotide
;   NAME/KEY: misc_feature
;   LOCATION: (0)...(0)
;   OTHER INFORMATION: phosphorothioate backbone
US-09-888-326-192

  Query Match             100.0%;  Score 8;  DB 9;  Length 20;
  Best Local Similarity   100.0%;  Pred. No. 1.5e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
            ||||||||
Db      11 CGCGCGCG 18


RESULT 4
US-09-888-326-192/c
; Sequence 192, Application US/09888326
; Publication No. US20030026801A1
; GENERAL INFORMATION:
;  APPLICANT: Weiner, George
;  APPLICANT:  Hartmann, Gunther
;  TITLE OF INVENTION: Methods for Enhancing Antibody-Induced
;  TITLE OF INVENTION:  Cell Lysis and Treating Cancer
;  FILE REFERENCE: C1039/7052 (AWS)
;  CURRENT APPLICATION NUMBER: US/09/888,326
;  CURRENT FILING DATE:  2001-06-22
;  PRIOR APPLICATION NUMBER: US 60/213,346
;  PRIOR FILING DATE: 2000-06-22
;  NUMBER OF SEQ ID NOS: 848
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 192
;   LENGTH: 20
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic oligonucleotide
;   NAME/KEY: misc_feature
;   LOCATION: (0)...(0)
;   OTHER INFORMATION: phosphorothioate backbone
US-09-888-326-192

  Query Match             100.0%;  Score 8;  DB 9;  Length 20;
  Best Local Similarity   100.0%;  Pred. No. 1.5e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
            ||||||||
Db      10 CGCGCGCG 3


RESULT 5
US-09-888-326-193
; Sequence 193, Application US/09888326
; Publication No. US20030026801A1
; GENERAL INFORMATION:
;  APPLICANT: Weiner, George
;  APPLICANT:  Hartmann, Gunther
;  TITLE OF INVENTION: Methods for Enhancing Antibody-Induced
;  TITLE OF INVENTION:  Cell Lysis and Treating Cancer
;  FILE REFERENCE: C1039/7052 (AWS)
;  CURRENT APPLICATION NUMBER: US/09/888,326
;  CURRENT FILING DATE:  2001-06-22
;  PRIOR APPLICATION NUMBER: US 60/213,346
;  PRIOR FILING DATE: 2000-06-22
;  NUMBER OF SEQ ID NOS: 848
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 193
;   LENGTH: 20
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic oligonucleotide
;   NAME/KEY: misc_feature
;   LOCATION: (0)...(0)
;   OTHER INFORMATION: phosphodiester backbone
US-09-888-326-193

  Query Match             100.0%;  Score 8;  DB 9;  Length 20;
  Best Local Similarity   100.0%;  Pred. No. 1.5e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
            ||||||||
Db      11 CGCGCGCG 18


RESULT 6
US-09-888-326-193/c
; Sequence 193, Application US/09888326
; Publication No. US20030026801A1
; GENERAL INFORMATION:
;  APPLICANT: Weiner, George
;  APPLICANT:  Hartmann, Gunther
;  TITLE OF INVENTION: Methods for Enhancing Antibody-Induced
```

```
;  TITLE OF INVENTION:  Cell Lysis and Treating Cancer
;  FILE REFERENCE: C1039/7052 (AWS)
;  CURRENT APPLICATION NUMBER: US/09/888,326
;  CURRENT FILING DATE:  2001-06-22
;  PRIOR APPLICATION NUMBER: US 60/213,346
;  PRIOR FILING DATE: 2000-06-22
;  NUMBER OF SEQ ID NOS: 848
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 193
;   LENGTH: 20
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic oligonucleotide
;   NAME/KEY: misc_feature
;   LOCATION: (0)...(0)
;   OTHER INFORMATION: phosphodiester backbone
US-09-888-326-193

  Query Match            100.0%;  Score 8;  DB 9;  Length 20;
  Best Local Similarity  100.0%;  Pred. No. 1.5e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
            ||||||||
Db      10 CGCGCGCG 3


RESULT 7
US-09-888-326-607
; Sequence 607, Application US/09888326
; Publication No. US20030026801A1
; GENERAL INFORMATION:
;  APPLICANT: Weiner, George
;  APPLICANT:  Hartmann, Gunther
;  TITLE OF INVENTION: Methods for Enhancing Antibody-Induced
;  TITLE OF INVENTION:  Cell Lysis and Treating Cancer
;  FILE REFERENCE: C1039/7052 (AWS)
;  CURRENT APPLICATION NUMBER: US/09/888,326
;  CURRENT FILING DATE:  2001-06-22
;  PRIOR APPLICATION NUMBER: US 60/213,346
;  PRIOR FILING DATE: 2000-06-22
;  NUMBER OF SEQ ID NOS: 848
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 607
;   LENGTH: 20
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic oligonucleotide
;   NAME/KEY: misc_feature
;   LOCATION: (0)...(0)
;   OTHER INFORMATION: phosphodiester backbone
US-09-888-326-607

  Query Match            100.0%;  Score 8;  DB 9;  Length 20;
  Best Local Similarity  100.0%;  Pred. No. 1.5e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
            ||||||||
Db      13 CGCGCGCG 20


RESULT 8
US-09-888-326-607/c
; Sequence 607, Application US/09888326
; Publication No. US20030026801A1
; GENERAL INFORMATION:
;  APPLICANT: Weiner, George
;  APPLICANT:  Hartmann, Gunther
;  TITLE OF INVENTION: Methods for Enhancing Antibody-Induced
;  TITLE OF INVENTION:  Cell Lysis and Treating Cancer
;  FILE REFERENCE: C1039/7052 (AWS)
;  CURRENT APPLICATION NUMBER: US/09/888,326
;  CURRENT FILING DATE:  2001-06-22
;  PRIOR APPLICATION NUMBER: US 60/213,346
;  PRIOR FILING DATE: 2000-06-22
;  NUMBER OF SEQ ID NOS: 848
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 607
;   LENGTH: 20
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic oligonucleotide
;   NAME/KEY: misc_feature
;   LOCATION: (0)...(0)
;   OTHER INFORMATION: phosphodiester backbone
US-09-888-326-607

  Query Match            100.0%;  Score 8;  DB 9;  Length 20;
  Best Local Similarity  100.0%;  Pred. No. 1.5e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
            ||||||||
Db      20 CGCGCGCG 13


RESULT 9
US-09-885-441-54
; Sequence 54, Application US/09885441
; Patent No. US20020146407A1
; GENERAL INFORMATION:
;  APPLICANT: Xiao, Yonghong
;  TITLE OF INVENTION: Regulation of Human Eosinophil Serine
;  TITLE OF INVENTION:  Protease-1-Like Enzyme
;  FILE REFERENCE: 04974.00512
;  CURRENT APPLICATION NUMBER: US/09/885,441
;  CURRENT FILING DATE:  2001-06-21
;  PRIOR APPLICATION NUMBER: US 60/212,844
;  PRIOR FILING DATE: 2000-06-21
;  PRIOR APPLICATION NUMBER: US 60/244,171
;  PRIOR FILING DATE: 2000-10-31
;  PRIOR APPLICATION NUMBER: US 60/279,766
;  PRIOR FILING DATE: 2001-03-30
;  PRIOR APPLICATION NUMBER: PCT/__________
;  PRIOR FILING DATE: 2001-06-20
;  NUMBER OF SEQ ID NOS: 58
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 54
;   LENGTH: 21
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-885-441-54

  Query Match            100.0%;  Score 8;  DB 10;  Length 21;
  Best Local Similarity  100.0%;  Pred. No. 1.5e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
            ||||||||
Db       6 CGCGCGCG 13


RESULT 10
US-09-885-441-54/c
; Sequence 54, Application US/09885441
; Patent No. US20020146407A1
; GENERAL INFORMATION:
;  APPLICANT: Xiao, Yonghong
;  TITLE OF INVENTION: Regulation of Human Eosinophil Serine
;  TITLE OF INVENTION:  Protease-1-Like Enzyme
;  FILE REFERENCE: 04974.00512
;  CURRENT APPLICATION NUMBER: US/09/885,441
```

```
;  CURRENT FILING DATE:  2001-06-21
;  PRIOR APPLICATION NUMBER: US 60/212,844
;  PRIOR FILING DATE: 2000-06-21
;  PRIOR APPLICATION NUMBER: US 60/244,171
;  PRIOR FILING DATE: 2000-10-31
;  PRIOR APPLICATION NUMBER: US 60/279,766
;  PRIOR FILING DATE: 2001-03-30
;  PRIOR APPLICATION NUMBER: PCT/_____
;  PRIOR FILING DATE: 2001-06-20
;  NUMBER OF SEQ ID NOS: 58
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 54
;   LENGTH: 21
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-885-441-54

  Query Match             100.0%;  Score 8;  DB 10;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 1.5e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
            ||||||||
Db      13 CGCGCGCG 6


RESULT 11
US-09-728-721-68
; Sequence 68, Application US/09728721
; Patent No. US20020061845A1
; GENERAL INFORMATION:
;  APPLICANT: Bertin, John
;  TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED PROTEIN FAMILY AND USES THERE
;  FILE REFERENCE: 07334-124001
;  CURRENT APPLICATION NUMBER: US/09/728,721
;  CURRENT FILING DATE:  2000-12-01
;  PRIOR APPLICATION NUMBER: 09/340,620
;  PRIOR FILING DATE: 1999-06-28
;  PRIOR APPLICATION NUMBER: US 09/207,359
;  PRIOR FILING DATE: 1998-12-08
;  PRIOR APPLICATION NUMBER: US 09/099,041
;  PRIOR FILING DATE: 1998-06-17
;  PRIOR APPLICATION NUMBER: US 09/019,942
;  PRIOR FILING DATE: 1998-02-06
;  NUMBER OF SEQ ID NOS: 71
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 68
;   LENGTH: 25
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-728-721-68

  Query Match             100.0%;  Score 8;  DB 10;  Length 25;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
            ||||||||
Db      16 CGCGCGCG 23


RESULT 12
US-09-728-721-68/c
; Sequence 68, Application US/09728721
; Patent No. US20020061845A1
; GENERAL INFORMATION:
;  APPLICANT: Bertin, John
;  TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED PROTEIN FAMILY AND USES THERE
;  FILE REFERENCE: 07334-124001
;  CURRENT APPLICATION NUMBER: US/09/728,721
;  CURRENT FILING DATE:  2000-12-01
;  PRIOR APPLICATION NUMBER: 09/340,620
;  PRIOR FILING DATE: 1999-06-28
;  PRIOR APPLICATION NUMBER: US 09/207,359
;  PRIOR FILING DATE: 1998-12-08
;  PRIOR APPLICATION NUMBER: US 09/099,041
;  PRIOR FILING DATE: 1998-06-17
;  PRIOR APPLICATION NUMBER: US 09/019,942
;  PRIOR FILING DATE: 1998-02-06
;  NUMBER OF SEQ ID NOS: 71
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 68
;   LENGTH: 25
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-728-721-68

  Query Match             100.0%;  Score 8;  DB 10;  Length 25;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
            ||||||||
Db      25 CGCGCGCG 18


RESULT 13
US-09-841-879B-16
; Sequence 16, Application US/09841879B
; Patent No. US20020142979A1
; GENERAL INFORMATION:
;  APPLICANT: Bertin, John
;  TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED PROTEIN FAMILY AND USES TH
;  FILE REFERENCE: 07334-330001
;  CURRENT APPLICATION NUMBER: US/09/841,879B
;  CURRENT FILING DATE:  2001-04-24
;  PRIOR APPLICATION NUMBER: US 09/728,721
;  PRIOR FILING DATE: 2000-12-01
;  PRIOR APPLICATION NUMBER: US 09/340,620
;  PRIOR FILING DATE: 1999-06-28
;  NUMBER OF SEQ ID NOS: 19
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16
;   LENGTH: 25
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-841-879B-16

  Query Match             100.0%;  Score 8;  DB 10;  Length 25;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
            ||||||||
Db      16 CGCGCGCG 23


RESULT 14
US-09-841-879B-16/c
; Sequence 16, Application US/09841879B
; Patent No. US20020142979A1
; GENERAL INFORMATION:
;  APPLICANT: Bertin, John
;  TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED PROTEIN FAMILY AND USES TH
;  FILE REFERENCE: 07334-330001
;  CURRENT APPLICATION NUMBER: US/09/841,879B
;  CURRENT FILING DATE:  2001-04-24
;  PRIOR APPLICATION NUMBER: US 09/728,721
;  PRIOR FILING DATE: 2000-12-01
;  PRIOR APPLICATION NUMBER: US 09/340,620
;  PRIOR FILING DATE: 1999-06-28
;  NUMBER OF SEQ ID NOS: 19
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16
;   LENGTH: 25
;   TYPE: DNA
```

```
;   ORGANISM: Homo sapiens
US-09-841-879B-16

  Query Match             100.0%;  Score 8;  DB 10;  Length 25;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db     25 CGCGCGCG 18


RESULT 15
US-09-682-597-4
; Sequence 4, Application US/09682597
; Patent No. US20020062503A1
; GENERAL INFORMATION:
;  APPLICANT: Monsanto Technology LLC
;  APPLICANT:  Chen, Guilan
;  APPLICANT:  Hironaka, Catherine
;  APPLICANT:  Zhou, Hua-Ping
;  TITLE OF INVENTION: Glyphosate Tolerant Wheat Plant 33391 and Compositions and Method
;  TITLE OF INVENTION:  Detection Thereof
;  FILE REFERENCE: 38-21(52232)A
;  CURRENT APPLICATION NUMBER: US/09/682,597
;  CURRENT FILING DATE:  2001-09-25
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
;   LENGTH: 30
;   TYPE: DNA
;   ORGANISM: Agrobacterium tumefaciens
US-09-682-597-4

  Query Match             100.0%;  Score 8;  DB 10;  Length 30;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db      3 CGCGCGCG 10


RESULT 16
US-09-682-597-4/c
; Sequence 4, Application US/09682597
; Patent No. US20020062503A1
; GENERAL INFORMATION:
;  APPLICANT: Monsanto Technology LLC
;  APPLICANT:  Chen, Guilan
;  APPLICANT:  Hironaka, Catherine
;  APPLICANT:  Zhou, Hua-Ping
;  TITLE OF INVENTION: Glyphosate Tolerant Wheat Plant 33391 and Compositions and Method
;  TITLE OF INVENTION:  Detection Thereof
;  FILE REFERENCE: 38-21(52232)A
;  CURRENT APPLICATION NUMBER: US/09/682,597
;  CURRENT FILING DATE:  2001-09-25
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
;   LENGTH: 30
;   TYPE: DNA
;   ORGANISM: Agrobacterium tumefaciens
US-09-682-597-4

  Query Match             100.0%;  Score 8;  DB 10;  Length 30;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db     10 CGCGCGCG 3


RESULT 17
US-09-828-034-5
; Sequence 5, Application US/09828034
; Patent No. US20020064771A1
; GENERAL INFORMATION:
;  APPLICANT: Zhong, Weidong
;  APPLICANT:  Hong, Zhi
;  APPLICANT:  Ferrari, Eric
;  TITLE OF INVENTION: HCV REPLICASE COMPLEXES
;  FILE REFERENCE: IN01165
;  CURRENT APPLICATION NUMBER: US/09/828,034
;  CURRENT FILING DATE:  2001-04-06
;  PRIOR APPLICATION NUMBER: U.S. 60/195,852
;  PRIOR FILING DATE: 2000-04-06
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
;   LENGTH: 36
;   TYPE: RNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic RNA
US-09-828-034-5

  Query Match             100.0%;  Score 8;  DB 10;  Length 36;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db     27 CGCGCGCG 34


RESULT 18
US-09-828-034-5/c
; Sequence 5, Application US/09828034
; Patent No. US20020064771A1
; GENERAL INFORMATION:
;  APPLICANT: Zhong, Weidong
;  APPLICANT:  Hong, Zhi
;  APPLICANT:  Ferrari, Eric
;  TITLE OF INVENTION: HCV REPLICASE COMPLEXES
;  FILE REFERENCE: IN01165
;  CURRENT APPLICATION NUMBER: US/09/828,034
;  CURRENT FILING DATE:  2001-04-06
;  PRIOR APPLICATION NUMBER: U.S. 60/195,852
;  PRIOR FILING DATE: 2000-04-06
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
;   LENGTH: 36
;   TYPE: RNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic RNA
US-09-828-034-5

  Query Match             100.0%;  Score 8;  DB 10;  Length 36;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db     36 CGCGCGCG 29


RESULT 19
US-09-733-042-31
; Sequence 31, Application US/09733042
; Patent No. US20020168709A1
; GENERAL INFORMATION:
;  APPLICANT:   Hennecke, Frank
```

```
;  APPLICANT:  Renner, Wolfgang A.
;  TITLE OF INVENTION:   Replicon Based Activation of Endogenous Genes
;  FILE REFERENCE:   1700.0100001
;  CURRENT APPLICATION NUMBER: US/09/733,042
;  CURRENT FILING DATE:  2000-12-11
;  PRIOR APPLICATION NUMBER:   US 60/169,988
;  PRIOR FILING DATE:   1999-12-10
;  NUMBER OF SEQ ID NOS:   49
;  SOFTWARE:   PatentIn version 3.0
; SEQ ID NO 31
;   LENGTH: 39
;   TYPE: DNA
;   ORGANISM: notsin5'-FOR
US-09-733-042-31

  Query Match             100.0%;  Score 8;  DB 9;  Length 39;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db      2 CGCGCGCG 9


RESULT 20
US-09-733-042-31/c
; Sequence 31, Application US/09733042
; Patent No. US20020168709A1
; GENERAL INFORMATION:
;  APPLICANT:   Hennecke, Frank
;  APPLICANT:  Renner, Wolfgang A.
;  TITLE OF INVENTION:   Replicon Based Activation of Endogenous Genes
;  FILE REFERENCE:   1700.0100001
;  CURRENT APPLICATION NUMBER: US/09/733,042
;  CURRENT FILING DATE:  2000-12-11
;  PRIOR APPLICATION NUMBER:   US 60/169,988
;  PRIOR FILING DATE:   1999-12-10
;  NUMBER OF SEQ ID NOS:   49
;  SOFTWARE:   PatentIn version 3.0
; SEQ ID NO 31
;   LENGTH: 39
;   TYPE: DNA
;   ORGANISM: notsin5'-FOR
US-09-733-042-31

  Query Match             100.0%;  Score 8;  DB 9;  Length 39;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db      9 CGCGCGCG 2


RESULT 21
US-10-007-132-45
; Sequence 45, Application US/10007132
; Publication No. US20030027254A1
;   GENERAL INFORMATION:
;        APPLICANT: Bard, Jonathan A
;                   Borowsky, Beth
;                   Smith, Kelli E
;        TITLE OF INVENTION: DNA ENCODING GALANIN GALR3 RECEPTORS
;                            AND USES THEREOF
;        NUMBER OF SEQUENCES: 65
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Cooper & Dunham LLP
;             STREET: 1185 Avenue of the Americas
;             CITY: New York
;             STATE: New York
;             COUNTRY: U.S.A.
;             ZIP: 10036
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/007,132
;             FILING DATE: 03-Dec-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 09/058,333
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: White, John P
;             REGISTRATION NUMBER: 28,678
;             REFERENCE/DOCKET NUMBER: 52241-E/JPW/KDB
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 212 278 0400
;             TELEFAX: 212 391 0525
;   INFORMATION FOR SEQ ID NO: 45:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 45 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: DNA
;        SEQUENCE DESCRIPTION: SEQ ID NO: 45:
US-10-007-132-45

  Query Match             100.0%;  Score 8;  DB 9;  Length 45;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db      7 CGCGCGCG 14


RESULT 22
US-10-007-132-45/c
; Sequence 45, Application US/10007132
; Publication No. US20030027254A1
;   GENERAL INFORMATION:
;        APPLICANT: Bard, Jonathan A
;                   Borowsky, Beth
;                   Smith, Kelli E
;        TITLE OF INVENTION: DNA ENCODING GALANIN GALR3 RECEPTORS
;                            AND USES THEREOF
;        NUMBER OF SEQUENCES: 65
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Cooper & Dunham LLP
;             STREET: 1185 Avenue of the Americas
;             CITY: New York
;             STATE: New York
;             COUNTRY: U.S.A.
;             ZIP: 10036
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/007,132
;             FILING DATE: 03-Dec-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 09/058,333
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: White, John P
;             REGISTRATION NUMBER: 28,678
;             REFERENCE/DOCKET NUMBER: 52241-E/JPW/KDB
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 212 278 0400
```

```
;            TELEFAX: 212 391 0525
;   INFORMATION FOR SEQ ID NO: 45:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 45 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: DNA
;        SEQUENCE DESCRIPTION: SEQ ID NO: 45:
US-10-007-132-45

  Query Match            100.0%;  Score 8;  DB 9;  Length 45;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db     14 CGCGCGCG 7


RESULT 23
US-09-790-417-19
; Sequence 19, Application US/09790417
; Patent No. US20010031470A1
; GENERAL INFORMATION:
;  APPLICANT: Shultz, John W
;  APPLICANT:  Lewis, Martin K.
;  APPLICANT:  Lieppe, Donna
;  APPLICANT:  Mandrekar, Michelle
;  APPLICANT:  Kephart, Daniel
;  APPLICANT:  Rhodes, Richard B.
;  APPLICANT:  Andrews, Christine A.
;  APPLICANT:  Hartnett, James R.
;  APPLICANT:  Gu, Trent
;  APPLICANT:  Olson, Ryan J.
;  APPLICANT:  Wood, Keith W.
;  APPLICANT:  Welch, Roy
;  TITLE OF INVENTION: Nucleic Acid Detection
;  FILE REFERENCE: Pro-103 6868/75528
;  CURRENT APPLICATION NUMBER: US/09/790,417
;  CURRENT FILING DATE:  2001-02-22
;  PRIOR APPLICATION NUMBER: 09/358,972
;  PRIOR FILING DATE: 1999-07-21
;  PRIOR APPLICATION NUMBER: 09/042,287
;  PRIOR FILING DATE: 1998-03-13
;  NUMBER OF SEQ ID NOS: 290
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
;   LENGTH: 55
;   TYPE: DNA
;   ORGANISM: Cytomegalovirus
;   FEATURE:
;   OTHER INFORMATION: target for mutant cytomegalovirus
US-09-790-417-19

  Query Match            100.0%;  Score 8;  DB 10;  Length 55;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db     27 CGCGCGCG 34


RESULT 24
US-09-790-417-19/c
; Sequence 19, Application US/09790417
; Patent No. US20010031470A1
; GENERAL INFORMATION:
;  APPLICANT: Shultz, John W
;  APPLICANT:  Lewis, Martin K.
;  APPLICANT:  Lieppe, Donna
;  APPLICANT:  Mandrekar, Michelle
;  APPLICANT:  Kephart, Daniel
;  APPLICANT:  Rhodes, Richard B.
;  APPLICANT:  Andrews, Christine A.
;  APPLICANT:  Hartnett, James R.
;  APPLICANT:  Gu, Trent
;  APPLICANT:  Olson, Ryan J.
;  APPLICANT:  Wood, Keith W.
;  APPLICANT:  Welch, Roy
;  TITLE OF INVENTION: Nucleic Acid Detection
;  FILE REFERENCE: Pro-103 6868/75528
;  CURRENT APPLICATION NUMBER: US/09/790,417
;  CURRENT FILING DATE:  2001-02-22
;  PRIOR APPLICATION NUMBER: 09/358,972
;  PRIOR FILING DATE: 1999-07-21
;  PRIOR APPLICATION NUMBER: 09/042,287
;  PRIOR FILING DATE: 1998-03-13
;  NUMBER OF SEQ ID NOS: 290
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
;   LENGTH: 55
;   TYPE: DNA
;   ORGANISM: Cytomegalovirus
;   FEATURE:
;   OTHER INFORMATION: target for mutant cytomegalovirus
US-09-790-417-19

  Query Match            100.0%;  Score 8;  DB 10;  Length 55;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db     34 CGCGCGCG 27


RESULT 25
US-09-899-381-37
; Sequence 37, Application US/09899381
; Patent No. US20020068293A1
; GENERAL INFORMATION:
;  APPLICANT: Delenstarr, Glend C.
;  APPLICANT:  Wolber, Pual K.
;  APPLICANT:  Sana, Theodore R.
;  TITLE OF INVENTION: Arrays Having Background Features and
;  TITLE OF INVENTION:  Methods for Using the Same
;  FILE REFERENCE: 10010760-1
;  CURRENT APPLICATION NUMBER: US/09/899,381
;  CURRENT FILING DATE:  2001-07-05
;  PRIOR APPLICATION NUMBER: 09/398,399
;  PRIOR FILING DATE: 1999-09-17
;  NUMBER OF SEQ ID NOS: 53
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 37
;   LENGTH: 60
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: synthetic probe
US-09-899-381-37

  Query Match            100.0%;  Score 8;  DB 10;  Length 60;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db     44 CGCGCGCG 51


RESULT 26
US-09-899-381-37/c
; Sequence 37, Application US/09899381
; Patent No. US20020068293A1
```

```
; GENERAL INFORMATION:
;  APPLICANT: Delenstarr, Glend C.
;  APPLICANT:  Wolber, Pual K.
;  APPLICANT:  Sana, Theodore R.
;  TITLE OF INVENTION: Arrays Having Background Features and
;  TITLE OF INVENTION:  Methods for Using the Same
;  FILE REFERENCE: 10010760-1
;  CURRENT APPLICATION NUMBER: US/09/899,381
;  CURRENT FILING DATE:  2001-07-05
;  PRIOR APPLICATION NUMBER: 09/398,399
;  PRIOR FILING DATE: 1999-09-17
;  NUMBER OF SEQ ID NOS: 53
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 37
;   LENGTH: 60
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: synthetic probe
US-09-899-381-37

  Query Match             100.0%;  Score 8;  DB 10;  Length 60;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db      17 CGCGCGCG 10


RESULT 27
US-10-014-973A-19
; Sequence 19, Application US/10014973A
; Patent No. US20020127581A1
; GENERAL INFORMATION:
;  APPLICANT: Ellington, Andrew
;  APPLICANT:  Jhaveri, Sulay
;  APPLICANT:  Rajendran, Manjula
;  TITLE OF INVENTION: In Vitro Selection of Signaling Aptamers
;  FILE REFERENCE: D6297
;  CURRENT APPLICATION NUMBER: US/10/014,973A
;  CURRENT FILING DATE:  2001-10-26
;  PRIOR APPLICATION NUMBER: US 60/244,010
;  PRIOR FILING DATE: 2000-10-27
;  NUMBER OF SEQ ID NOS: 19
; SEQ ID NO 19
;   LENGTH: 61
;   TYPE: RNA
;   ORGANISM: artificial sequence
;   FEATURE:
;   LOCATION: 48
;   OTHER INFORMATION: Sequence of the cloned aptamer raf17s in
;   OTHER INFORMATION: Family 1; u is 12-F-u at position 48
US-10-014-973A-19

  Query Match             100.0%;  Score 8;  DB 12;  Length 61;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db      51 CGCGCGCG 58


RESULT 28
US-10-014-973A-19/c
; Sequence 19, Application US/10014973A
; Patent No. US20020127581A1
; GENERAL INFORMATION:
;  APPLICANT: Ellington, Andrew
;  APPLICANT:  Jhaveri, Sulay
;  APPLICANT:  Rajendran, Manjula
;  TITLE OF INVENTION: In Vitro Selection of Signaling Aptamers
;  FILE REFERENCE: D6297
;  CURRENT APPLICATION NUMBER: US/10/014,973A
;  CURRENT FILING DATE:  2001-10-26
;  PRIOR APPLICATION NUMBER: US 60/244,010
;  PRIOR FILING DATE: 2000-10-27
;  NUMBER OF SEQ ID NOS: 19
; SEQ ID NO 19
;   LENGTH: 61
;   TYPE: RNA
;   ORGANISM: artificial sequence
;   FEATURE:
;   LOCATION: 48
;   OTHER INFORMATION: Sequence of the cloned aptamer raf17s in
;   OTHER INFORMATION: Family 1; u is 12-F-u at position 48
US-10-014-973A-19

  Query Match             100.0%;  Score 8;  DB 12;  Length 61;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db      58 CGCGCGCG 51


RESULT 29
US-10-057-940-8
; Sequence 8, Application US/10057940
; Patent No. US20020168686A1
; GENERAL INFORMATION:
;  APPLICANT: Pantoliano, Michael W.
;  APPLICANT:  Salemme, F. Raymond
;  APPLICANT:  Carver, Jr., Theodore, E.
;  TITLE OF INVENTION: High Throughput Method for Functionally Classifying Proteins
;  TITLE OF INVENTION:  Identified Using a Genomics Approach
;  FILE REFERENCE: 1503.0310002/JAG/JSO
;  CURRENT APPLICATION NUMBER: US/10/057,940
;  CURRENT FILING DATE:  2002-05-03
;  PRIOR APPLICATION NUMBER: 09/190,128
;  PRIOR FILING DATE: 1998-11-12
;  PRIOR APPLICATION NUMBER: US 60/065,129
;  PRIOR FILING DATE: 1997-11-12
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
;   LENGTH: 64
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: GC-rich tract
US-10-057-940-8

  Query Match             100.0%;  Score 8;  DB 9;  Length 64;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db      55 CGCGCGCG 62


RESULT 30
US-10-057-940-8/c
; Sequence 8, Application US/10057940
; Patent No. US20020168686A1
; GENERAL INFORMATION:
;  APPLICANT: Pantoliano, Michael W.
;  APPLICANT:  Salemme, F. Raymond
;  APPLICANT:  Carver, Jr., Theodore, E.
;  TITLE OF INVENTION: High Throughput Method for Functionally Classifying Proteins
;  TITLE OF INVENTION:  Identified Using a Genomics Approach
;  FILE REFERENCE: 1503.0310002/JAG/JSO
;  CURRENT APPLICATION NUMBER: US/10/057,940
```

```
;  CURRENT FILING DATE:  2002-05-03
;  PRIOR APPLICATION NUMBER: 09/190,128
;  PRIOR FILING DATE: 1998-11-12
;  PRIOR APPLICATION NUMBER: US 60/065,129
;  PRIOR FILING DATE: 1997-11-12
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
;   LENGTH: 64
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: GC-rich tract
US-10-057-940-8

  Query Match           100.0%;  Score 8;  DB 9;  Length 64;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db      9 CGCGCGCG 2


RESULT 31
US-09-473-872-54
; Sequence 54, Application US/09473872
; Patent No. US20020064876A1
; GENERAL INFORMATION:
;  APPLICANT:  Yoon, Kyonggeun
;  TITLE OF INVENTION:  No. US20020064876A1el Gene Therapy Methods For The Treatment of
;  TITLE OF INVENTION:  Disorders
;  FILE REFERENCE:  JEFF-YOO01
;  CURRENT APPLICATION NUMBER: US/09/473,872
;  CURRENT FILING DATE:  1999-12-28
;  NUMBER OF SEQ ID NOS:  68
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 54
;   LENGTH: 68
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Combined DNA/RNA Molecule:
;   OTHER INFORMATION: Chimeric oligonucleotide
;   OTHER INFORMATION: Description of Artificial Sequence:  Chimeric
;   OTHER INFORMATION: oligonucleotide
US-09-473-872-54

  Query Match           100.0%;  Score 8;  DB 10;  Length 68;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db     52 CGCGCGCG 59


RESULT 32
US-09-473-872-54/c
; Sequence 54, Application US/09473872
; Patent No. US20020064876A1
; GENERAL INFORMATION:
;  APPLICANT:  Yoon, Kyonggeun
;  TITLE OF INVENTION:  No. US20020064876A1el Gene Therapy Methods For The Treatment of
;  TITLE OF INVENTION:  Disorders
;  FILE REFERENCE:  JEFF-YOO01
;  CURRENT APPLICATION NUMBER: US/09/473,872
;  CURRENT FILING DATE:  1999-12-28
;  NUMBER OF SEQ ID NOS:  68
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 54
;   LENGTH: 68
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Combined DNA/RNA Molecule:
;   OTHER INFORMATION: Chimeric oligonucleotide
;   OTHER INFORMATION: Description of Artificial Sequence:  Chimeric
;   OTHER INFORMATION: oligonucleotide
US-09-473-872-54

  Query Match           100.0%;  Score 8;  DB 10;  Length 68;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db     59 CGCGCGCG 52


RESULT 33
US-09-835-976B-122
; Sequence 122, Application US/09835976B
; Publication No. US20030027983A1
; GENERAL INFORMATION:
;  APPLICANT: Mount, David B.
;  APPLICANT:  Delpire, Eric
;  APPLICANT:  Gamba, Gerardo
;  APPLICANT:  Alfred L. George, Jr.
;  TITLE OF INVENTION: PURIFIED AND ISOLATED POTASSIUM-CHLORIDE COTRANSPORTER NUCLEIC
;  TITLE OF INVENTION:  POLYPEPTIDES AND
;  TITLE OF INVENTION:  THERAPEUTIC AND SCREENING METHODS USING SAME
;  FILE REFERENCE: Attorney Docket No. US20030027983A1 1242-26-2
;  CURRENT APPLICATION NUMBER: US/09/835,976B
;  CURRENT FILING DATE:  2001-04-16
;  NUMBER OF SEQ ID NOS: 131
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 122
;   LENGTH: 72
;   TYPE: DNA
;   ORGANISM: homo sapiens
US-09-835-976B-122

  Query Match           100.0%;  Score 8;  DB 9;  Length 72;
  Best Local Similarity   100.0%;  Pred. No. 1e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db     30 CGCGCGCG 37


RESULT 34
US-09-835-976B-122/c
; Sequence 122, Application US/09835976B
; Publication No. US20030027983A1
; GENERAL INFORMATION:
;  APPLICANT: Mount, David B.
;  APPLICANT:  Delpire, Eric
;  APPLICANT:  Gamba, Gerardo
;  APPLICANT:  Alfred L. George, Jr.
;  TITLE OF INVENTION: PURIFIED AND ISOLATED POTASSIUM-CHLORIDE COTRANSPORTER NUCLEIC
;  TITLE OF INVENTION:  POLYPEPTIDES AND
;  TITLE OF INVENTION:  THERAPEUTIC AND SCREENING METHODS USING SAME
;  FILE REFERENCE: Attorney Docket No. US20030027983A1 1242-26-2
;  CURRENT APPLICATION NUMBER: US/09/835,976B
;  CURRENT FILING DATE:  2001-04-16
;  NUMBER OF SEQ ID NOS: 131
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 122
;   LENGTH: 72
;   TYPE: DNA
;   ORGANISM: homo sapiens
US-09-835-976B-122

  Query Match           100.0%;  Score 8;  DB 9;  Length 72;
```

```
  Best Local Similarity   100.0%;  Pred. No. 1e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db     37 CGCGCGCG 30


RESULT 35
US-09-835-976B-119
; Sequence 119, Application US/09835976B
; Publication No. US20030027983A1
; GENERAL INFORMATION:
;  APPLICANT: Mount, David B.
;  APPLICANT:  Delpire, Eric
;  APPLICANT:  Gamba, Gerardo
;  APPLICANT:  Alfred L. George, Jr.
;  TITLE OF INVENTION: PURIFIED AND ISOLATED POTASSIUM-CHLORIDE COTRANSPORTER NUCLEIC AC
;  TITLE OF INVENTION:  POLYPEPTIDES AND
;  TITLE OF INVENTION:  THERAPEUTIC AND SCREENING METHODS USING SAME
;  FILE REFERENCE: Attorney Docket No. US20030027983A1 1242-26-2
;  CURRENT APPLICATION NUMBER: US/09/835,976B
;  CURRENT FILING DATE:  2001-04-16
;  NUMBER OF SEQ ID NOS: 131
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 119
;   LENGTH: 74
;   TYPE: DNA
;   ORGANISM: homo sapiens
US-09-835-976B-119

  Query Match             100.0%;  Score 8;  DB 9;  Length 74;
  Best Local Similarity   100.0%;  Pred. No. 1e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db     34 CGCGCGCG 41


RESULT 36
US-09-835-976B-119/c
; Sequence 119, Application US/09835976B
; Publication No. US20030027983A1
; GENERAL INFORMATION:
;  APPLICANT: Mount, David B.
;  APPLICANT:  Delpire, Eric
;  APPLICANT:  Gamba, Gerardo
;  APPLICANT:  Alfred L. George, Jr.
;  TITLE OF INVENTION: PURIFIED AND ISOLATED POTASSIUM-CHLORIDE COTRANSPORTER NUCLEIC AC
;  TITLE OF INVENTION:  POLYPEPTIDES AND
;  TITLE OF INVENTION:  THERAPEUTIC AND SCREENING METHODS USING SAME
;  FILE REFERENCE: Attorney Docket No. US20030027983A1 1242-26-2
;  CURRENT APPLICATION NUMBER: US/09/835,976B
;  CURRENT FILING DATE:  2001-04-16
;  NUMBER OF SEQ ID NOS: 131
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 119
;   LENGTH: 74
;   TYPE: DNA
;   ORGANISM: homo sapiens
US-09-835-976B-119

  Query Match             100.0%;  Score 8;  DB 9;  Length 74;
  Best Local Similarity   100.0%;  Pred. No. 1e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db     41 CGCGCGCG 34


RESULT 37
```

```
US-09-835-976B-127
; Sequence 127, Application US/09835976B
; Publication No. US20030027983A1
; GENERAL INFORMATION:
;  APPLICANT: Mount, David B.
;  APPLICANT:  Delpire, Eric
;  APPLICANT:  Gamba, Gerardo
;  APPLICANT:  Alfred L. George, Jr.
;  TITLE OF INVENTION: PURIFIED AND ISOLATED POTASSIUM-CHLORIDE COTRANSPORTER NUCLEIC
;  TITLE OF INVENTION:  POLYPEPTIDES AND
;  TITLE OF INVENTION:  THERAPEUTIC AND SCREENING METHODS USING SAME
;  FILE REFERENCE: Attorney Docket No. US20030027983A1 1242-26-2
;  CURRENT APPLICATION NUMBER: US/09/835,976B
;  CURRENT FILING DATE:  2001-04-16
;  NUMBER OF SEQ ID NOS: 131
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 127
;   LENGTH: 76
;   TYPE: DNA
;   ORGANISM: homo sapiens
US-09-835-976B-127

  Query Match             100.0%;  Score 8;  DB 9;  Length 76;
  Best Local Similarity   100.0%;  Pred. No. 1e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db     32 CGCGCGCG 39


RESULT 38
US-09-835-976B-127/c
; Sequence 127, Application US/09835976B
; Publication No. US20030027983A1
; GENERAL INFORMATION:
;  APPLICANT: Mount, David B.
;  APPLICANT:  Delpire, Eric
;  APPLICANT:  Gamba, Gerardo
;  APPLICANT:  Alfred L. George, Jr.
;  TITLE OF INVENTION: PURIFIED AND ISOLATED POTASSIUM-CHLORIDE COTRANSPORTER NUCLEIC
;  TITLE OF INVENTION:  POLYPEPTIDES AND
;  TITLE OF INVENTION:  THERAPEUTIC AND SCREENING METHODS USING SAME
;  FILE REFERENCE: Attorney Docket No. US20030027983A1 1242-26-2
;  CURRENT APPLICATION NUMBER: US/09/835,976B
;  CURRENT FILING DATE:  2001-04-16
;  NUMBER OF SEQ ID NOS: 131
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 127
;   LENGTH: 76
;   TYPE: DNA
;   ORGANISM: homo sapiens
US-09-835-976B-127

  Query Match             100.0%;  Score 8;  DB 9;  Length 76;
  Best Local Similarity   100.0%;  Pred. No. 1e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db     41 CGCGCGCG 34


RESULT 39
US-09-835-976B-129
; Sequence 129, Application US/09835976B
; Publication No. US20030027983A1
; GENERAL INFORMATION:
;  APPLICANT: Mount, David B.
;  APPLICANT:  Delpire, Eric
;  APPLICANT:  Gamba, Gerardo
;  APPLICANT:  Alfred L. George, Jr.
;  TITLE OF INVENTION: PURIFIED AND ISOLATED POTASSIUM-CHLORIDE COTRANSPORTER NUCLEIC
```

```
;  TITLE OF INVENTION:  POLYPEPTIDES AND
;  TITLE OF INVENTION:  THERAPEUTIC AND SCREENING METHODS USING SAME
;  FILE REFERENCE: Attorney Docket No. US20030027983A1 1242-26-2
;  CURRENT APPLICATION NUMBER: US/09/835,976B
;  CURRENT FILING DATE:  2001-04-16
;  NUMBER OF SEQ ID NOS: 131
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 129
;   LENGTH: 76
;   TYPE: DNA
;   ORGANISM: homo sapiens
US-09-835-976B-129

  Query Match             100.0%;  Score 8;  DB 9;  Length 76;
  Best Local Similarity   100.0%;  Pred. No. 1e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
            ||||||||
Db      34 CGCGCGCG 41


RESULT 40
US-09-835-976B-129/c
; Sequence 129, Application US/09835976B
; Publication No. US20030027983A1
; GENERAL INFORMATION:
;  APPLICANT: Mount, David B.
;  APPLICANT:  Delpire, Eric
;  APPLICANT:  Gamba, Gerardo
;  APPLICANT:  Alfred L. George, Jr.
;  TITLE OF INVENTION: PURIFIED AND ISOLATED POTASSIUM-CHLORIDE COTRANSPORTER NUCLEIC AC
;  TITLE OF INVENTION:  POLYPEPTIDES AND
;  TITLE OF INVENTION:  THERAPEUTIC AND SCREENING METHODS USING SAME
;  FILE REFERENCE: Attorney Docket No. US20030027983A1 1242-26-2
;  CURRENT APPLICATION NUMBER: US/09/835,976B
;  CURRENT FILING DATE:  2001-04-16
;  NUMBER OF SEQ ID NOS: 131
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 129
;   LENGTH: 76
;   TYPE: DNA
;   ORGANISM: homo sapiens
US-09-835-976B-129

  Query Match             100.0%;  Score 8;  DB 9;  Length 76;
  Best Local Similarity   100.0%;  Pred. No. 1e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
            ||||||||
Db      41 CGCGCGCG 34


RESULT 41
US-09-835-976B-126
; Sequence 126, Application US/09835976B
; Publication No. US20030027983A1
; GENERAL INFORMATION:
;  APPLICANT: Mount, David B.
;  APPLICANT:  Delpire, Eric
;  APPLICANT:  Gamba, Gerardo
;  APPLICANT:  Alfred L. George, Jr.
;  TITLE OF INVENTION: PURIFIED AND ISOLATED POTASSIUM-CHLORIDE COTRANSPORTER NUCLEIC AC
;  TITLE OF INVENTION:  POLYPEPTIDES AND
;  TITLE OF INVENTION:  THERAPEUTIC AND SCREENING METHODS USING SAME
;  FILE REFERENCE: Attorney Docket No. US20030027983A1 1242-26-2
;  CURRENT APPLICATION NUMBER: US/09/835,976B
;  CURRENT FILING DATE:  2001-04-16
;  NUMBER OF SEQ ID NOS: 131
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 126
;   LENGTH: 78
;   TYPE: DNA
;   ORGANISM: homo sapiens
US-09-835-976B-126

  Query Match             100.0%;  Score 8;  DB 9;  Length 78;
  Best Local Similarity   100.0%;  Pred. No. 1e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
            ||||||||
Db      34 CGCGCGCG 41


RESULT 42
US-09-835-976B-126/c
; Sequence 126, Application US/09835976B
; Publication No. US20030027983A1
; GENERAL INFORMATION:
;  APPLICANT: Mount, David B.
;  APPLICANT:  Delpire, Eric
;  APPLICANT:  Gamba, Gerardo
;  APPLICANT:  Alfred L. George, Jr.
;  TITLE OF INVENTION: PURIFIED AND ISOLATED POTASSIUM-CHLORIDE COTRANSPORTER NUCLEIC
;  TITLE OF INVENTION:  POLYPEPTIDES AND
;  TITLE OF INVENTION:  THERAPEUTIC AND SCREENING METHODS USING SAME
;  FILE REFERENCE: Attorney Docket No. US20030027983A1 1242-26-2
;  CURRENT APPLICATION NUMBER: US/09/835,976B
;  CURRENT FILING DATE:  2001-04-16
;  NUMBER OF SEQ ID NOS: 131
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 126
;   LENGTH: 78
;   TYPE: DNA
;   ORGANISM: homo sapiens
US-09-835-976B-126

  Query Match             100.0%;  Score 8;  DB 9;  Length 78;
  Best Local Similarity   100.0%;  Pred. No. 1e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
            ||||||||
Db      43 CGCGCGCG 36


RESULT 43
US-09-835-976B-128
; Sequence 128, Application US/09835976B
; Publication No. US20030027983A1
; GENERAL INFORMATION:
;  APPLICANT: Mount, David B.
;  APPLICANT:  Delpire, Eric
;  APPLICANT:  Gamba, Gerardo
;  APPLICANT:  Alfred L. George, Jr.
;  TITLE OF INVENTION: PURIFIED AND ISOLATED POTASSIUM-CHLORIDE COTRANSPORTER NUCLEIC
;  TITLE OF INVENTION:  POLYPEPTIDES AND
;  TITLE OF INVENTION:  THERAPEUTIC AND SCREENING METHODS USING SAME
;  FILE REFERENCE: Attorney Docket No. US20030027983A1 1242-26-2
;  CURRENT APPLICATION NUMBER: US/09/835,976B
;  CURRENT FILING DATE:  2001-04-16
;  NUMBER OF SEQ ID NOS: 131
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 128
;   LENGTH: 78
;   TYPE: DNA
;   ORGANISM: homo sapiens
US-09-835-976B-128

  Query Match             100.0%;  Score 8;  DB 9;  Length 78;
  Best Local Similarity   100.0%;  Pred. No. 1e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
```

```
              ||||||||
Db      34 CGCGCGCG 41


RESULT 44
US-09-835-976B-128/c
; Sequence 128, Application US/09835976B
; Publication No. US20030027983A1
; GENERAL INFORMATION:
;  APPLICANT: Mount, David B.
;  APPLICANT:  Delpire, Eric
;  APPLICANT:  Gamba, Gerardo
;  APPLICANT:  Alfred L. George, Jr.
;  TITLE OF INVENTION: PURIFIED AND ISOLATED POTASSIUM-CHLORIDE COTRANSPORTER NUCLEIC AC
;  TITLE OF INVENTION:  POLYPEPTIDES AND
;  TITLE OF INVENTION:  THERAPEUTIC AND SCREENING METHODS USING SAME
;  FILE REFERENCE: Attorney Docket No. US20030027983A1 1242-26-2
;  CURRENT APPLICATION NUMBER: US/09/835,976B
;  CURRENT FILING DATE:  2001-04-16
;  NUMBER OF SEQ ID NOS: 131
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 128
;   LENGTH: 78
;   TYPE: DNA
;   ORGANISM: homo sapiens
US-09-835-976B-128

  Query Match             100.0%;  Score 8;  DB 9;  Length 78;
  Best Local Similarity   100.0%;  Pred. No. 1e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
              ||||||||
Db      43 CGCGCGCG 36


RESULT 45
US-09-835-976B-124
; Sequence 124, Application US/09835976B
; Publication No. US20030027983A1
; GENERAL INFORMATION:
;  APPLICANT: Mount, David B.
;  APPLICANT:  Delpire, Eric
;  APPLICANT:  Gamba, Gerardo
;  APPLICANT:  Alfred L. George, Jr.
;  TITLE OF INVENTION: PURIFIED AND ISOLATED POTASSIUM-CHLORIDE COTRANSPORTER NUCLEIC AC
;  TITLE OF INVENTION:  POLYPEPTIDES AND
;  TITLE OF INVENTION:  THERAPEUTIC AND SCREENING METHODS USING SAME
;  FILE REFERENCE: Attorney Docket No. US20030027983A1 1242-26-2
;  CURRENT APPLICATION NUMBER: US/09/835,976B
;  CURRENT FILING DATE:  2001-04-16
;  NUMBER OF SEQ ID NOS: 131
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 124
;   LENGTH: 80
;   TYPE: DNA
;   ORGANISM: homo sapiens
US-09-835-976B-124

  Query Match             100.0%;  Score 8;  DB 9;  Length 80;
  Best Local Similarity   100.0%;  Pred. No. 1e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
              ||||||||
Db      36 CGCGCGCG 43


Search completed: March 14, 2003, 04:34:58
Job time : 65 secs
```

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: March 14, 2003, 03:37:53 ; Search time 43 Seconds
(without alignments)
57.056 Million cell updates/sec

Title: CGCGCGCG
Perfect score: 8
Sequence: 1 cgcgcgcg 8

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	8	100.0	10	2	US-08-734-973-9	Sequence 9, Appli
c	2	8	100.0	10	2	US-08-734-973-9	Sequence 9, Appli
	3	8	100.0	10	2	US-08-734-973-10	Sequence 10, Appl
c	4	8	100.0	10	2	US-08-734-973-10	Sequence 10, Appl
	5	8	100.0	10	2	US-08-734-973-11	Sequence 11, Appl
c	6	8	100.0	10	2	US-08-734-973-11	Sequence 11, Appl
	7	8	100.0	10	2	US-08-734-973-12	Sequence 12, Appl
c	8	8	100.0	10	2	US-08-734-973-12	Sequence 12, Appl
	9	8	100.0	10	2	US-08-734-973-13	Sequence 13, Appl
c	10	8	100.0	10	2	US-08-734-973-13	Sequence 13, Appl
	11	8	100.0	10	2	US-08-734-973-14	Sequence 14, Appl
c	12	8	100.0	10	2	US-08-734-973-14	Sequence 14, Appl
	13	8	100.0	10	2	US-08-734-973-15	Sequence 15, Appl
c	14	8	100.0	10	2	US-08-734-973-15	Sequence 15, Appl
	15	8	100.0	10	2	US-08-734-973-16	Sequence 16, Appl
c	16	8	100.0	10	2	US-08-734-973-16	Sequence 16, Appl
	17	8	100.0	10	2	US-08-734-973-17	Sequence 17, Appl
c	18	8	100.0	10	2	US-08-734-973-17	Sequence 17, Appl
	19	8	100.0	10	2	US-08-734-973-18	Sequence 18, Appl
c	20	8	100.0	10	2	US-08-734-973-18	Sequence 18, Appl
	21	8	100.0	10	2	US-08-734-973-19	Sequence 19, Appl
c	22	8	100.0	10	2	US-08-734-973-19	Sequence 19, Appl
	23	8	100.0	10	2	US-08-734-973-20	Sequence 20, Appl
c	24	8	100.0	10	2	US-08-734-973-20	Sequence 20, Appl
	25	8	100.0	10	2	US-08-734-973-21	Sequence 21, Appl
c	26	8	100.0	10	2	US-08-734-973-21	Sequence 21, Appl
	27	8	100.0	10	2	US-08-734-973-22	Sequence 22, Appl
c	28	8	100.0	10	2	US-08-734-973-22	Sequence 22, Appl
	29	8	100.0	10	2	US-08-734-973-23	Sequence 23, Appl
c	30	8	100.0	10	2	US-08-734-973-23	Sequence 23, Appl
	31	8	100.0	10	2	US-08-734-973-24	Sequence 24, Appl
c	32	8	100.0	10	2	US-08-734-973-24	Sequence 24, Appl
	33	8	100.0	10	2	US-08-734-973-25	Sequence 25, Appl
c	34	8	100.0	10	2	US-08-734-973-25	Sequence 25, Appl
	35	8	100.0	10	2	US-08-734-973-26	Sequence 26, Appl
c	36	8	100.0	10	2	US-08-734-973-26	Sequence 26, Appl
	37	8	100.0	10	2	US-08-734-973-27	Sequence 27, Appl
c	38	8	100.0	10	2	US-08-734-973-27	Sequence 27, Appl
	39	8	100.0	10	3	US-08-729-598-13	Sequence 13, Appl
c	40	8	100.0	10	3	US-08-729-598-13	Sequence 13, Appl
	41	8	100.0	12	4	US-09-393-783A-79	Sequence 79, Appl
c	42	8	100.0	12	4	US-09-393-783A-79	Sequence 79, Appl
	43	8	100.0	12	4	US-09-151-890B-79	Sequence 79, Appl
c	44	8	100.0	12	4	US-09-151-890B-79	Sequence 79, Appl
	45	8	100.0	14	2	US-08-595-043A-44	Sequence 44, Appl

ALIGNMENTS

```
RESULT 1
US-08-734-973-9
; Sequence 9, Application US/08734973
; Patent No. 5912147
;  GENERAL INFORMATION:
;    APPLICANT:  Stoler, Daniel L.
;    APPLICANT:  Basik, Mark
;    APPLICANT:  Anderson, Garth R.
;    TITLE OF INVENTION:  A Rapid Means For Quantitating
;    TITLE OF INVENTION:  Genomic Instability
;    NUMBER OF SEQUENCES:  38
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hodgson, Russ, Andrews, Woods & Goodyear
;      STREET:  1800 One M&T Plaza
;      CITY:  Buffalo
;      STATE:  New York
;      COUNTRY:  United States
;      ZIP:  14203-2391
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS/ Microsoft Windows
;      SOFTWARE:  Wordperfect for Windows
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/734,973
;      FILING DATE:  October 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Nelson, M. Bud
;      REGISTRATION NUMBER:  35,300
;      REFERENCE/DOCKET NUMBER:  03551.0021
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (716) 856-4000
;      TELEFAX:  (716) 849-0349
;  INFORMATION FOR SEQ ID NO:  9 :
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10 nucleotides
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single-stranded
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
;    HYPOTHETICAL:  No
US-08-734-973-9

  Query Match           100.0%;  Score 8;  DB 2;  Length 10;
  Best Local Similarity  100.0%;  Pred. No. 1.2e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       1 CGCGCGCG 8
```

```
RESULT 2
US-08-734-973-9/c
; Sequence 9, Application US/08734973
; Patent No. 5912147
;  GENERAL INFORMATION:
;    APPLICANT:  Stoler, Daniel L.
;    APPLICANT:  Basik, Mark
;    APPLICANT:  Anderson, Garth R.
;    TITLE OF INVENTION:  A Rapid Means For Quantitating
;    TITLE OF INVENTION:  Genomic Instability
;    NUMBER OF SEQUENCES:  38
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hodgson, Russ, Andrews, Woods & Goodyear
;      STREET:  1800 One M&T Plaza
;      CITY:  Buffalo
;      STATE:  New York
;      COUNTRY:  United States
;      ZIP:  14203-2391
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS/ Microsoft Windows
;      SOFTWARE:  Wordperfect for Windows
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/734,973
;      FILING DATE:  October 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Nelson, M. Bud
;      REGISTRATION NUMBER:  35,300
;      REFERENCE/DOCKET NUMBER:  03551.0021
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (716) 856-4000
;      TELEFAX:  (716) 849-0349
;  INFORMATION FOR SEQ ID NO:  9 :
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10 nucleotides
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single-stranded
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
;    HYPOTHETICAL:  No
US-08-734-973-9

  Query Match             100.0%;  Score 8;  DB 2;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       8 CGCGCGCG 1


RESULT 3
US-08-734-973-10
; Sequence 10, Application US/08734973
; Patent No. 5912147
;  GENERAL INFORMATION:
;    APPLICANT:  Stoler, Daniel L.
;    APPLICANT:  Basik, Mark
;    APPLICANT:  Anderson, Garth R.
;    TITLE OF INVENTION:  A Rapid Means For Quantitating
;    TITLE OF INVENTION:  Genomic Instability
;    NUMBER OF SEQUENCES:  38
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hodgson, Russ, Andrews, Woods & Goodyear
;      STREET:  1800 One M&T Plaza
;      CITY:  Buffalo
;      STATE:  New York
;      COUNTRY:  United States
;      ZIP:  14203-2391
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS/ Microsoft Windows
;      SOFTWARE:  Wordperfect for Windows
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/734,973
;      FILING DATE:  October 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Nelson, M. Bud
;      REGISTRATION NUMBER:  35,300
;      REFERENCE/DOCKET NUMBER:  03551.0021
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (716) 856-4000
;      TELEFAX:  (716) 849-0349
;  INFORMATION FOR SEQ ID NO:  10 :
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10 nucleotides
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single-stranded
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
;    HYPOTHETICAL:  No
US-08-734-973-10

  Query Match             100.0%;  Score 8;  DB 2;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       1 CGCGCGCG 8


RESULT 4
US-08-734-973-10/c
; Sequence 10, Application US/08734973
; Patent No. 5912147
;  GENERAL INFORMATION:
;    APPLICANT:  Stoler, Daniel L.
;    APPLICANT:  Basik, Mark
;    APPLICANT:  Anderson, Garth R.
;    TITLE OF INVENTION:  A Rapid Means For Quantitating
;    TITLE OF INVENTION:  Genomic Instability
;    NUMBER OF SEQUENCES:  38
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hodgson, Russ, Andrews, Woods & Goodyear
;      STREET:  1800 One M&T Plaza
;      CITY:  Buffalo
;      STATE:  New York
;      COUNTRY:  United States
;      ZIP:  14203-2391
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS/ Microsoft Windows
;      SOFTWARE:  Wordperfect for Windows
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/734,973
;      FILING DATE:  October 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Nelson, M. Bud
;      REGISTRATION NUMBER:  35,300
;      REFERENCE/DOCKET NUMBER:  03551.0021
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (716) 856-4000
;      TELEFAX:  (716) 849-0349
;  INFORMATION FOR SEQ ID NO:  10 :
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10 nucleotides
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single-stranded
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
```

```
;    HYPOTHETICAL:  No
US-08-734-973-10

  Query Match             100.0%;  Score 8;  DB 2;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db      8 CGCGCGCG 1


RESULT 5
US-08-734-973-11
; Sequence 11, Application US/08734973
; Patent No. 5912147
;  GENERAL INFORMATION:
;    APPLICANT:  Stoler, Daniel L.
;    APPLICANT:  Basik, Mark
;    APPLICANT:  Anderson, Garth R.
;    TITLE OF INVENTION:  A Rapid Means For Quantitating
;    TITLE OF INVENTION:  Genomic Instability
;    NUMBER OF SEQUENCES:  38
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hodgson, Russ, Andrews, Woods & Goodyear
;      STREET:  1800 One M&T Plaza
;      CITY:  Buffalo
;      STATE:  New York
;      COUNTRY:  United States
;      ZIP:  14203-2391
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS/ Microsoft Windows
;      SOFTWARE:  Wordperfect for Windows
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/734,973
;      FILING DATE:  October 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Nelson, M. Bud
;      REGISTRATION NUMBER:  35,300
;      REFERENCE/DOCKET NUMBER:  03551.0021
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (716) 856-4000
;      TELEFAX:  (716) 849-0349
;  INFORMATION FOR SEQ ID NO:  11 :
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10 nucleotides
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single-stranded
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
;    HYPOTHETICAL:  No
US-08-734-973-11

  Query Match             100.0%;  Score 8;  DB 2;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db      1 CGCGCGCG 8


RESULT 6
US-08-734-973-11/c
; Sequence 11, Application US/08734973
; Patent No. 5912147
;  GENERAL INFORMATION:
;    APPLICANT:  Stoler, Daniel L.
;    APPLICANT:  Basik, Mark
;    APPLICANT:  Anderson, Garth R.
;    TITLE OF INVENTION:  A Rapid Means For Quantitating
;    TITLE OF INVENTION:  Genomic Instability
;    NUMBER OF SEQUENCES:  38
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hodgson, Russ, Andrews, Woods & Goodyear
;      STREET:  1800 One M&T Plaza
;      CITY:  Buffalo
;      STATE:  New York
;      COUNTRY:  United States
;      ZIP:  14203-2391
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS/ Microsoft Windows
;      SOFTWARE:  Wordperfect for Windows
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/734,973
;      FILING DATE:  October 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Nelson, M. Bud
;      REGISTRATION NUMBER:  35,300
;      REFERENCE/DOCKET NUMBER:  03551.0021
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (716) 856-4000
;      TELEFAX:  (716) 849-0349
;  INFORMATION FOR SEQ ID NO:  11 :
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10 nucleotides
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single-stranded
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
;    HYPOTHETICAL:  No
US-08-734-973-11

  Query Match             100.0%;  Score 8;  DB 2;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db      8 CGCGCGCG 1


RESULT 7
US-08-734-973-12
; Sequence 12, Application US/08734973
; Patent No. 5912147
;  GENERAL INFORMATION:
;    APPLICANT:  Stoler, Daniel L.
;    APPLICANT:  Basik, Mark
;    APPLICANT:  Anderson, Garth R.
;    TITLE OF INVENTION:  A Rapid Means For Quantitating
;    TITLE OF INVENTION:  Genomic Instability
;    NUMBER OF SEQUENCES:  38
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hodgson, Russ, Andrews, Woods & Goodyear
;      STREET:  1800 One M&T Plaza
;      CITY:  Buffalo
;      STATE:  New York
;      COUNTRY:  United States
;      ZIP:  14203-2391
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS/ Microsoft Windows
;      SOFTWARE:  Wordperfect for Windows
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/734,973
;      FILING DATE:  October 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Nelson, M. Bud
;      REGISTRATION NUMBER:  35,300
;      REFERENCE/DOCKET NUMBER:  03551.0021
```

```
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (716) 856-4000
;      TELEFAX:  (716) 849-0349
;  INFORMATION FOR SEQ ID NO:  12 :
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10 nucleotides
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single-stranded
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
;    HYPOTHETICAL:  No
US-08-734-973-12

  Query Match           100.0%;  Score 8;  DB 2;  Length 10;
  Best Local Similarity  100.0%;  Pred. No. 1.2e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db      1 CGCGCGCG 8


RESULT 8
US-08-734-973-12/c
; Sequence 12, Application US/08734973
; Patent No. 5912147
;  GENERAL INFORMATION:
;    APPLICANT:  Stoler, Daniel L.
;    APPLICANT:  Basik, Mark
;    APPLICANT:  Anderson, Garth R.
;    TITLE OF INVENTION:  A Rapid Means For Quantitating
;    TITLE OF INVENTION:  Genomic Instability
;    NUMBER OF SEQUENCES:  38
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hodgson, Russ, Andrews, Woods & Goodyear
;      STREET:  1800 One M&T Plaza
;      CITY:  Buffalo
;      STATE:  New York
;      COUNTRY:  United States
;      ZIP:  14203-2391
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS/ Microsoft Windows
;      SOFTWARE:  Wordperfect for Windows
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/734,973
;      FILING DATE:  October 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Nelson, M. Bud
;      REGISTRATION NUMBER:  35,300
;      REFERENCE/DOCKET NUMBER:  03551.0021
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (716) 856-4000
;      TELEFAX:  (716) 849-0349
;  INFORMATION FOR SEQ ID NO:  12 :
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10 nucleotides
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single-stranded
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
;    HYPOTHETICAL:  No
US-08-734-973-12

  Query Match           100.0%;  Score 8;  DB 2;  Length 10;
  Best Local Similarity  100.0%;  Pred. No. 1.2e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db      8 CGCGCGCG 1


RESULT 9
US-08-734-973-13
; Sequence 13, Application US/08734973
; Patent No. 5912147
;  GENERAL INFORMATION:
;    APPLICANT:  Stoler, Daniel L.
;    APPLICANT:  Basik, Mark
;    APPLICANT:  Anderson, Garth R.
;    TITLE OF INVENTION:  A Rapid Means For Quantitating
;    TITLE OF INVENTION:  Genomic Instability
;    NUMBER OF SEQUENCES:  38
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hodgson, Russ, Andrews, Woods & Goodyear
;      STREET:  1800 One M&T Plaza
;      CITY:  Buffalo
;      STATE:  New York
;      COUNTRY:  United States
;      ZIP:  14203-2391
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS/ Microsoft Windows
;      SOFTWARE:  Wordperfect for Windows
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/734,973
;      FILING DATE:  October 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Nelson, M. Bud
;      REGISTRATION NUMBER:  35,300
;      REFERENCE/DOCKET NUMBER:  03551.0021
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (716) 856-4000
;      TELEFAX:  (716) 849-0349
;  INFORMATION FOR SEQ ID NO:  13 :
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10 nucleotides
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single-stranded
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
;    HYPOTHETICAL:  No
US-08-734-973-13

  Query Match           100.0%;  Score 8;  DB 2;  Length 10;
  Best Local Similarity  100.0%;  Pred. No. 1.2e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db      1 CGCGCGCG 8


RESULT 10
US-08-734-973-13/c
; Sequence 13, Application US/08734973
; Patent No. 5912147
;  GENERAL INFORMATION:
;    APPLICANT:  Stoler, Daniel L.
;    APPLICANT:  Basik, Mark
;    APPLICANT:  Anderson, Garth R.
;    TITLE OF INVENTION:  A Rapid Means For Quantitating
;    TITLE OF INVENTION:  Genomic Instability
;    NUMBER OF SEQUENCES:  38
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hodgson, Russ, Andrews, Woods & Goodyear
;      STREET:  1800 One M&T Plaza
;      CITY:  Buffalo
;      STATE:  New York
;      COUNTRY:  United States
;      ZIP:  14203-2391
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch
```

```
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  MS-DOS/ Microsoft Windows
;       SOFTWARE:  Wordperfect for Windows
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/734,973
;       FILING DATE:  October 1996
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Nelson, M. Bud
;       REGISTRATION NUMBER:  35,300
;       REFERENCE/DOCKET NUMBER:  03551.0021
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (716) 856-4000
;       TELEFAX:  (716) 849-0349
;   INFORMATION FOR SEQ ID NO:  13 :
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  10 nucleotides
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single-stranded
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA
;     HYPOTHETICAL:  No
US-08-734-973-13

  Query Match             100.0%;  Score 8;  DB 2;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       8 CGCGCGCG 1


RESULT 11
US-08-734-973-14
; Sequence 14, Application US/08734973
; Patent No. 5912147
;  GENERAL INFORMATION:
;    APPLICANT:  Stoler, Daniel L.
;    APPLICANT:  Basik, Mark
;    APPLICANT:  Anderson, Garth R.
;    TITLE OF INVENTION:  A Rapid Means For Quantitating
;    TITLE OF INVENTION:  Genomic Instability
;    NUMBER OF SEQUENCES:  38
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hodgson, Russ, Andrews, Woods & Goodyear
;      STREET:  1800 One M&T Plaza
;      CITY:  Buffalo
;      STATE:  New York
;      COUNTRY:  United States
;      ZIP:  14203-2391
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS/ Microsoft Windows
;      SOFTWARE:  Wordperfect for Windows
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/734,973
;      FILING DATE:  October 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Nelson, M. Bud
;      REGISTRATION NUMBER:  35,300
;      REFERENCE/DOCKET NUMBER:  03551.0021
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (716) 856-4000
;      TELEFAX:  (716) 849-0349
;  INFORMATION FOR SEQ ID NO:  14 :
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10 nucleotides
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single-stranded
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
;    HYPOTHETICAL:  No
US-08-734-973-14

  Query Match             100.0%;  Score 8;  DB 2;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       1 CGCGCGCG 8


RESULT 12
US-08-734-973-14/c
; Sequence 14, Application US/08734973
; Patent No. 5912147
;  GENERAL INFORMATION:
;    APPLICANT:  Stoler, Daniel L.
;    APPLICANT:  Basik, Mark
;    APPLICANT:  Anderson, Garth R.
;    TITLE OF INVENTION:  A Rapid Means For Quantitating
;    TITLE OF INVENTION:  Genomic Instability
;    NUMBER OF SEQUENCES:  38
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hodgson, Russ, Andrews, Woods & Goodyear
;      STREET:  1800 One M&T Plaza
;      CITY:  Buffalo
;      STATE:  New York
;      COUNTRY:  United States
;      ZIP:  14203-2391
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS/ Microsoft Windows
;      SOFTWARE:  Wordperfect for Windows
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/734,973
;      FILING DATE:  October 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Nelson, M. Bud
;      REGISTRATION NUMBER:  35,300
;      REFERENCE/DOCKET NUMBER:  03551.0021
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (716) 856-4000
;      TELEFAX:  (716) 849-0349
;  INFORMATION FOR SEQ ID NO:  14 :
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10 nucleotides
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single-stranded
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
;    HYPOTHETICAL:  No
US-08-734-973-14

  Query Match             100.0%;  Score 8;  DB 2;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       8 CGCGCGCG 1


RESULT 13
US-08-734-973-15
; Sequence 15, Application US/08734973
; Patent No. 5912147
;  GENERAL INFORMATION:
;    APPLICANT:  Stoler, Daniel L.
;    APPLICANT:  Basik, Mark
;    APPLICANT:  Anderson, Garth R.
;    TITLE OF INVENTION:  A Rapid Means For Quantitating
;    TITLE OF INVENTION:  Genomic Instability
```

```
;    NUMBER OF SEQUENCES:  38
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hodgson, Russ, Andrews, Woods & Goodyear
;      STREET:  1800 One M&T Plaza
;      CITY:  Buffalo
;      STATE:  New York
;      COUNTRY:  United States
;      ZIP:  14203-2391
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS/ Microsoft Windows
;      SOFTWARE:  Wordperfect for Windows
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/734,973
;      FILING DATE:  October 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Nelson, M. Bud
;      REGISTRATION NUMBER:  35,300
;      REFERENCE/DOCKET NUMBER:  03551.0021
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (716) 856-4000
;      TELEFAX:  (716) 849-0349
;  INFORMATION FOR SEQ ID NO:  15 :
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10 nucleotides
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single-stranded
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
;    HYPOTHETICAL:  No
US-08-734-973-15

  Query Match             100.0%;  Score 8;  DB 2;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       1 CGCGCGCG 8


RESULT 14
US-08-734-973-15/c
; Sequence 15, Application US/08734973
; Patent No. 5912147
;  GENERAL INFORMATION:
;    APPLICANT:  Stoler, Daniel L.
;    APPLICANT:  Basik, Mark
;    APPLICANT:  Anderson, Garth R.
;    TITLE OF INVENTION:  A Rapid Means For Quantitating
;    TITLE OF INVENTION:  Genomic Instability
;    NUMBER OF SEQUENCES:  38
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hodgson, Russ, Andrews, Woods & Goodyear
;      STREET:  1800 One M&T Plaza
;      CITY:  Buffalo
;      STATE:  New York
;      COUNTRY:  United States
;      ZIP:  14203-2391
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS/ Microsoft Windows
;      SOFTWARE:  Wordperfect for Windows
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/734,973
;      FILING DATE:  October 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Nelson, M. Bud
;      REGISTRATION NUMBER:  35,300
;      REFERENCE/DOCKET NUMBER:  03551.0021
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (716) 856-4000
;      TELEFAX:  (716) 849-0349
;  INFORMATION FOR SEQ ID NO:  15 :
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10 nucleotides
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single-stranded
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
;    HYPOTHETICAL:  No
US-08-734-973-15

  Query Match             100.0%;  Score 8;  DB 2;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       8 CGCGCGCG 1


RESULT 15
US-08-734-973-16
; Sequence 16, Application US/08734973
; Patent No. 5912147
;  GENERAL INFORMATION:
;    APPLICANT:  Stoler, Daniel L.
;    APPLICANT:  Basik, Mark
;    APPLICANT:  Anderson, Garth R.
;    TITLE OF INVENTION:  A Rapid Means For Quantitating
;    TITLE OF INVENTION:  Genomic Instability
;    NUMBER OF SEQUENCES:  38
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hodgson, Russ, Andrews, Woods & Goodyear
;      STREET:  1800 One M&T Plaza
;      CITY:  Buffalo
;      STATE:  New York
;      COUNTRY:  United States
;      ZIP:  14203-2391
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS/ Microsoft Windows
;      SOFTWARE:  Wordperfect for Windows
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/734,973
;      FILING DATE:  October 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Nelson, M. Bud
;      REGISTRATION NUMBER:  35,300
;      REFERENCE/DOCKET NUMBER:  03551.0021
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (716) 856-4000
;      TELEFAX:  (716) 849-0349
;  INFORMATION FOR SEQ ID NO:  16 :
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10 nucleotides
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single-stranded
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
;    HYPOTHETICAL:  No
US-08-734-973-16

  Query Match             100.0%;  Score 8;  DB 2;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       1 CGCGCGCG 8
```

```
RESULT 16
US-08-734-973-16/c
; Sequence 16, Application US/08734973
; Patent No. 5912147
;  GENERAL INFORMATION:
;    APPLICANT:  Stoler, Daniel L.
;    APPLICANT:  Basik, Mark
;    APPLICANT:  Anderson, Garth R.
;    TITLE OF INVENTION:  A Rapid Means For Quantitating
;    TITLE OF INVENTION:  Genomic Instability
;    NUMBER OF SEQUENCES:  38
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hodgson, Russ, Andrews, Woods & Goodyear
;      STREET:  1800 One M&T Plaza
;      CITY:  Buffalo
;      STATE:  New York
;      COUNTRY:  United States
;      ZIP:  14203-2391
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS/ Microsoft Windows
;      SOFTWARE:  Wordperfect for Windows
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/734,973
;      FILING DATE:  October 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Nelson, M. Bud
;      REGISTRATION NUMBER:  35,300
;      REFERENCE/DOCKET NUMBER:  03551.0021
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (716) 856-4000
;      TELEFAX:  (716) 849-0349
;  INFORMATION FOR SEQ ID NO:  16 :
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10 nucleotides
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single-stranded
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
;    HYPOTHETICAL:  No
US-08-734-973-16

  Query Match             100.0%;  Score 8;  DB 2;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       8 CGCGCGCG 1


RESULT 17
US-08-734-973-17
; Sequence 17, Application US/08734973
; Patent No. 5912147
;  GENERAL INFORMATION:
;    APPLICANT:  Stoler, Daniel L.
;    APPLICANT:  Basik, Mark
;    APPLICANT:  Anderson, Garth R.
;    TITLE OF INVENTION:  A Rapid Means For Quantitating
;    TITLE OF INVENTION:  Genomic Instability
;    NUMBER OF SEQUENCES:  38
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hodgson, Russ, Andrews, Woods & Goodyear
;      STREET:  1800 One M&T Plaza
;      CITY:  Buffalo
;      STATE:  New York
;      COUNTRY:  United States
;      ZIP:  14203-2391
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS/ Microsoft Windows
;      SOFTWARE:  Wordperfect for Windows
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/734,973
;      FILING DATE:  October 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Nelson, M. Bud
;      REGISTRATION NUMBER:  35,300
;      REFERENCE/DOCKET NUMBER:  03551.0021
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (716) 856-4000
;      TELEFAX:  (716) 849-0349
;  INFORMATION FOR SEQ ID NO:  17 :
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10 nucleotides
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single-stranded
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
;    HYPOTHETICAL:  No
US-08-734-973-17

  Query Match             100.0%;  Score 8;  DB 2;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       1 CGCGCGCG 8


RESULT 18
US-08-734-973-17/c
; Sequence 17, Application US/08734973
; Patent No. 5912147
;  GENERAL INFORMATION:
;    APPLICANT:  Stoler, Daniel L.
;    APPLICANT:  Basik, Mark
;    APPLICANT:  Anderson, Garth R.
;    TITLE OF INVENTION:  A Rapid Means For Quantitating
;    TITLE OF INVENTION:  Genomic Instability
;    NUMBER OF SEQUENCES:  38
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hodgson, Russ, Andrews, Woods & Goodyear
;      STREET:  1800 One M&T Plaza
;      CITY:  Buffalo
;      STATE:  New York
;      COUNTRY:  United States
;      ZIP:  14203-2391
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS/ Microsoft Windows
;      SOFTWARE:  Wordperfect for Windows
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/734,973
;      FILING DATE:  October 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Nelson, M. Bud
;      REGISTRATION NUMBER:  35,300
;      REFERENCE/DOCKET NUMBER:  03551.0021
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (716) 856-4000
;      TELEFAX:  (716) 849-0349
;  INFORMATION FOR SEQ ID NO:  17 :
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10 nucleotides
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single-stranded
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
;    HYPOTHETICAL:  No
US-08-734-973-17
```

```
  Query Match             100.0%;  Score 8;  DB 2;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       8 CGCGCGCG 1


RESULT 19
US-08-734-973-18
; Sequence 18, Application US/08734973
; Patent No. 5912147
;  GENERAL INFORMATION:
;    APPLICANT:  Stoler, Daniel L.
;    APPLICANT:  Basik, Mark
;    APPLICANT:  Anderson, Garth R.
;    TITLE OF INVENTION:  A Rapid Means For Quantitating
;    TITLE OF INVENTION:  Genomic Instability
;    NUMBER OF SEQUENCES:  38
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hodgson, Russ, Andrews, Woods & Goodyear
;      STREET:  1800 One M&T Plaza
;      CITY:  Buffalo
;      STATE:  New York
;      COUNTRY:  United States
;      ZIP:  14203-2391
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS/ Microsoft Windows
;      SOFTWARE:  Wordperfect for Windows
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/734,973
;      FILING DATE:  October 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Nelson, M. Bud
;      REGISTRATION NUMBER:  35,300
;      REFERENCE/DOCKET NUMBER:  03551.0021
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (716) 856-4000
;      TELEFAX:  (716) 849-0349
;  INFORMATION FOR SEQ ID NO:  18 :
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10 nucleotides
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single-stranded
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
;    HYPOTHETICAL:  No
US-08-734-973-18

  Query Match             100.0%;  Score 8;  DB 2;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       1 CGCGCGCG 8


RESULT 20
US-08-734-973-18/c
; Sequence 18, Application US/08734973
; Patent No. 5912147
;  GENERAL INFORMATION:
;    APPLICANT:  Stoler, Daniel L.
;    APPLICANT:  Basik, Mark
;    APPLICANT:  Anderson, Garth R.
;    TITLE OF INVENTION:  A Rapid Means For Quantitating
;    TITLE OF INVENTION:  Genomic Instability
;    NUMBER OF SEQUENCES:  38
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hodgson, Russ, Andrews, Woods & Goodyear
;      STREET:  1800 One M&T Plaza
;      CITY:  Buffalo
;      STATE:  New York
;      COUNTRY:  United States
;      ZIP:  14203-2391
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS/ Microsoft Windows
;      SOFTWARE:  Wordperfect for Windows
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/734,973
;      FILING DATE:  October 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Nelson, M. Bud
;      REGISTRATION NUMBER:  35,300
;      REFERENCE/DOCKET NUMBER:  03551.0021
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (716) 856-4000
;      TELEFAX:  (716) 849-0349
;  INFORMATION FOR SEQ ID NO:  18 :
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10 nucleotides
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single-stranded
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
;    HYPOTHETICAL:  No
US-08-734-973-18

  Query Match             100.0%;  Score 8;  DB 2;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       8 CGCGCGCG 1


RESULT 21
US-08-734-973-19
; Sequence 19, Application US/08734973
; Patent No. 5912147
;  GENERAL INFORMATION:
;    APPLICANT:  Stoler, Daniel L.
;    APPLICANT:  Basik, Mark
;    APPLICANT:  Anderson, Garth R.
;    TITLE OF INVENTION:  A Rapid Means For Quantitating
;    TITLE OF INVENTION:  Genomic Instability
;    NUMBER OF SEQUENCES:  38
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hodgson, Russ, Andrews, Woods & Goodyear
;      STREET:  1800 One M&T Plaza
;      CITY:  Buffalo
;      STATE:  New York
;      COUNTRY:  United States
;      ZIP:  14203-2391
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS/ Microsoft Windows
;      SOFTWARE:  Wordperfect for Windows
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/734,973
;      FILING DATE:  October 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Nelson, M. Bud
;      REGISTRATION NUMBER:  35,300
;      REFERENCE/DOCKET NUMBER:  03551.0021
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (716) 856-4000
```

```
;       TELEFAX:  (716) 849-0349
;   INFORMATION FOR SEQ ID NO:  19 :
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  10 nucleotides
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single-stranded
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA
;     HYPOTHETICAL:  No
US-08-734-973-19

  Query Match             100.0%;  Score 8;  DB 2;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       1 CGCGCGCG 8


RESULT 22
US-08-734-973-19/c
; Sequence 19, Application US/08734973
; Patent No. 5912147
;  GENERAL INFORMATION:
;     APPLICANT:  Stoler, Daniel L.
;     APPLICANT:  Basik, Mark
;     APPLICANT:  Anderson, Garth R.
;     TITLE OF INVENTION:  A Rapid Means For Quantitating
;     TITLE OF INVENTION:  Genomic Instability
;     NUMBER OF SEQUENCES:  38
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Hodgson, Russ, Andrews, Woods & Goodyear
;       STREET:  1800 One M&T Plaza
;       CITY:  Buffalo
;       STATE:  New York
;       COUNTRY:  United States
;       ZIP:  14203-2391
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette, 3.5 inch
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  MS-DOS/ Microsoft Windows
;       SOFTWARE:  Wordperfect for Windows
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/734,973
;       FILING DATE:  October 1996
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Nelson, M. Bud
;       REGISTRATION NUMBER:  35,300
;       REFERENCE/DOCKET NUMBER:  03551.0021
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (716) 856-4000
;       TELEFAX:  (716) 849-0349
;   INFORMATION FOR SEQ ID NO:  19 :
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  10 nucleotides
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single-stranded
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA
;     HYPOTHETICAL:  No
US-08-734-973-19

  Query Match             100.0%;  Score 8;  DB 2;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       8 CGCGCGCG 1


RESULT 23
US-08-734-973-20
; Sequence 20, Application US/08734973
; Patent No. 5912147
;  GENERAL INFORMATION:
;     APPLICANT:  Stoler, Daniel L.
;     APPLICANT:  Basik, Mark
;     APPLICANT:  Anderson, Garth R.
;     TITLE OF INVENTION:  A Rapid Means For Quantitating
;     TITLE OF INVENTION:  Genomic Instability
;     NUMBER OF SEQUENCES:  38
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Hodgson, Russ, Andrews, Woods & Goodyear
;       STREET:  1800 One M&T Plaza
;       CITY:  Buffalo
;       STATE:  New York
;       COUNTRY:  United States
;       ZIP:  14203-2391
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette, 3.5 inch
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  MS-DOS/ Microsoft Windows
;       SOFTWARE:  Wordperfect for Windows
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/734,973
;       FILING DATE:  October 1996
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Nelson, M. Bud
;       REGISTRATION NUMBER:  35,300
;       REFERENCE/DOCKET NUMBER:  03551.0021
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (716) 856-4000
;       TELEFAX:  (716) 849-0349
;   INFORMATION FOR SEQ ID NO:  20 :
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  10 nucleotides
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single-stranded
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA
;     HYPOTHETICAL:  No
US-08-734-973-20

  Query Match             100.0%;  Score 8;  DB 2;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       1 CGCGCGCG 8


RESULT 24
US-08-734-973-20/c
; Sequence 20, Application US/08734973
; Patent No. 5912147
;  GENERAL INFORMATION:
;     APPLICANT:  Stoler, Daniel L.
;     APPLICANT:  Basik, Mark
;     APPLICANT:  Anderson, Garth R.
;     TITLE OF INVENTION:  A Rapid Means For Quantitating
;     TITLE OF INVENTION:  Genomic Instability
;     NUMBER OF SEQUENCES:  38
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Hodgson, Russ, Andrews, Woods & Goodyear
;       STREET:  1800 One M&T Plaza
;       CITY:  Buffalo
;       STATE:  New York
;       COUNTRY:  United States
;       ZIP:  14203-2391
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette, 3.5 inch
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  MS-DOS/ Microsoft Windows
```

```
;       SOFTWARE:  Wordperfect for Windows
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/734,973
;       FILING DATE:  October 1996
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Nelson, M. Bud
;       REGISTRATION NUMBER:  35,300
;       REFERENCE/DOCKET NUMBER:  03551.0021
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (716) 856-4000
;       TELEFAX:  (716) 849-0349
;   INFORMATION FOR SEQ ID NO:  20 :
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  10 nucleotides
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single-stranded
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA
;     HYPOTHETICAL:  No
US-08-734-973-20

  Query Match             100.0%;  Score 8;  DB 2;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       8 CGCGCGCG 1


RESULT 25
US-08-734-973-21
; Sequence 21, Application US/08734973
; Patent No. 5912147
;  GENERAL INFORMATION:
;     APPLICANT:  Stoler, Daniel L.
;     APPLICANT:  Basik, Mark
;     APPLICANT:  Anderson, Garth R.
;     TITLE OF INVENTION:  A Rapid Means For Quantitating
;     TITLE OF INVENTION:  Genomic Instability
;     NUMBER OF SEQUENCES:  38
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Hodgson, Russ, Andrews, Woods & Goodyear
;       STREET:  1800 One M&T Plaza
;       CITY:  Buffalo
;       STATE:  New York
;       COUNTRY:  United States
;       ZIP:  14203-2391
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette, 3.5 inch
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  MS-DOS/ Microsoft Windows
;       SOFTWARE:  Wordperfect for Windows
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/734,973
;       FILING DATE:  October 1996
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Nelson, M. Bud
;       REGISTRATION NUMBER:  35,300
;       REFERENCE/DOCKET NUMBER:  03551.0021
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (716) 856-4000
;       TELEFAX:  (716) 849-0349
;   INFORMATION FOR SEQ ID NO:  21 :
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  10 nucleotides
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single-stranded
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA
;     HYPOTHETICAL:  No
US-08-734-973-21

  Query Match             100.0%;  Score 8;  DB 2;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       1 CGCGCGCG 8


RESULT 26
US-08-734-973-21/c
; Sequence 21, Application US/08734973
; Patent No. 5912147
;  GENERAL INFORMATION:
;     APPLICANT:  Stoler, Daniel L.
;     APPLICANT:  Basik, Mark
;     APPLICANT:  Anderson, Garth R.
;     TITLE OF INVENTION:  A Rapid Means For Quantitating
;     TITLE OF INVENTION:  Genomic Instability
;     NUMBER OF SEQUENCES:  38
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Hodgson, Russ, Andrews, Woods & Goodyear
;       STREET:  1800 One M&T Plaza
;       CITY:  Buffalo
;       STATE:  New York
;       COUNTRY:  United States
;       ZIP:  14203-2391
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette, 3.5 inch
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  MS-DOS/ Microsoft Windows
;       SOFTWARE:  Wordperfect for Windows
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/734,973
;       FILING DATE:  October 1996
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Nelson, M. Bud
;       REGISTRATION NUMBER:  35,300
;       REFERENCE/DOCKET NUMBER:  03551.0021
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (716) 856-4000
;       TELEFAX:  (716) 849-0349
;   INFORMATION FOR SEQ ID NO:  21 :
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  10 nucleotides
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single-stranded
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA
;     HYPOTHETICAL:  No
US-08-734-973-21

  Query Match             100.0%;  Score 8;  DB 2;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       8 CGCGCGCG 1


RESULT 27
US-08-734-973-22
; Sequence 22, Application US/08734973
; Patent No. 5912147
;  GENERAL INFORMATION:
;     APPLICANT:  Stoler, Daniel L.
;     APPLICANT:  Basik, Mark
;     APPLICANT:  Anderson, Garth R.
;     TITLE OF INVENTION:  A Rapid Means For Quantitating
;     TITLE OF INVENTION:  Genomic Instability
;     NUMBER OF SEQUENCES:  38
;     CORRESPONDENCE ADDRESS:
```

```
;       ADDRESSEE:  Hodgson, Russ, Andrews, Woods & Goodyear
;       STREET:  1800 One M&T Plaza
;       CITY:  Buffalo
;       STATE:  New York
;       COUNTRY:  United States
;       ZIP:  14203-2391
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette, 3.5 inch
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  MS-DOS/ Microsoft Windows
;       SOFTWARE:  Wordperfect for Windows
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/734,973
;       FILING DATE:  October 1996
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Nelson, M. Bud
;       REGISTRATION NUMBER:  35,300
;       REFERENCE/DOCKET NUMBER:  03551.0021
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (716) 856-4000
;       TELEFAX:  (716) 849-0349
;   INFORMATION FOR SEQ ID NO:  22 :
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  10 nucleotides
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single-stranded
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA
;     HYPOTHETICAL:  No
US-08-734-973-22

  Query Match             100.0%;  Score 8;  DB 2;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       1 CGCGCGCG 8


RESULT 28
US-08-734-973-22/c
; Sequence 22, Application US/08734973
; Patent No. 5912147
;  GENERAL INFORMATION:
;    APPLICANT:  Stoler, Daniel L.
;    APPLICANT:  Basik, Mark
;    APPLICANT:  Anderson, Garth R.
;    TITLE OF INVENTION:  A Rapid Means For Quantitating
;    TITLE OF INVENTION:  Genomic Instability
;    NUMBER OF SEQUENCES:  38
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hodgson, Russ, Andrews, Woods & Goodyear
;      STREET:  1800 One M&T Plaza
;      CITY:  Buffalo
;      STATE:  New York
;      COUNTRY:  United States
;      ZIP:  14203-2391
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS/ Microsoft Windows
;      SOFTWARE:  Wordperfect for Windows
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/734,973
;      FILING DATE:  October 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Nelson, M. Bud
;      REGISTRATION NUMBER:  35,300
;      REFERENCE/DOCKET NUMBER:  03551.0021
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (716) 856-4000
;      TELEFAX:  (716) 849-0349
;  INFORMATION FOR SEQ ID NO:  22 :
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10 nucleotides
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single-stranded
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
;    HYPOTHETICAL:  No
US-08-734-973-22

  Query Match             100.0%;  Score 8;  DB 2;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       8 CGCGCGCG 1


RESULT 29
US-08-734-973-23
; Sequence 23, Application US/08734973
; Patent No. 5912147
;  GENERAL INFORMATION:
;    APPLICANT:  Stoler, Daniel L.
;    APPLICANT:  Basik, Mark
;    APPLICANT:  Anderson, Garth R.
;    TITLE OF INVENTION:  A Rapid Means For Quantitating
;    TITLE OF INVENTION:  Genomic Instability
;    NUMBER OF SEQUENCES:  38
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hodgson, Russ, Andrews, Woods & Goodyear
;      STREET:  1800 One M&T Plaza
;      CITY:  Buffalo
;      STATE:  New York
;      COUNTRY:  United States
;      ZIP:  14203-2391
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS/ Microsoft Windows
;      SOFTWARE:  Wordperfect for Windows
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/734,973
;      FILING DATE:  October 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Nelson, M. Bud
;      REGISTRATION NUMBER:  35,300
;      REFERENCE/DOCKET NUMBER:  03551.0021
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (716) 856-4000
;      TELEFAX:  (716) 849-0349
;  INFORMATION FOR SEQ ID NO:  23 :
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10 nucleotides
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single-stranded
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
;    HYPOTHETICAL:  No
US-08-734-973-23

  Query Match             100.0%;  Score 8;  DB 2;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       1 CGCGCGCG 8


RESULT 30
US-08-734-973-23/c
```

```
; Sequence 23, Application US/08734973
; Patent No. 5912147
;  GENERAL INFORMATION:
;    APPLICANT:  Stoler, Daniel L.
;    APPLICANT:  Basik, Mark
;    APPLICANT:  Anderson, Garth R.
;    TITLE OF INVENTION:  A Rapid Means For Quantitating
;    TITLE OF INVENTION:  Genomic Instability
;    NUMBER OF SEQUENCES:  38
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hodgson, Russ, Andrews, Woods & Goodyear
;      STREET:  1800 One M&T Plaza
;      CITY:  Buffalo
;      STATE:  New York
;      COUNTRY:  United States
;      ZIP:  14203-2391
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS/ Microsoft Windows
;      SOFTWARE:  Wordperfect for Windows
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/734,973
;      FILING DATE:  October 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Nelson, M. Bud
;      REGISTRATION NUMBER:  35,300
;      REFERENCE/DOCKET NUMBER:  03551.0021
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (716) 856-4000
;      TELEFAX:  (716) 849-0349
;  INFORMATION FOR SEQ ID NO:  23 :
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10 nucleotides
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single-stranded
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
;    HYPOTHETICAL:  No
US-08-734-973-23

  Query Match             100.0%;  Score 8;  DB 2;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       8 CGCGCGCG 1


RESULT 31
US-08-734-973-24
; Sequence 24, Application US/08734973
; Patent No. 5912147
;  GENERAL INFORMATION:
;    APPLICANT:  Stoler, Daniel L.
;    APPLICANT:  Basik, Mark
;    APPLICANT:  Anderson, Garth R.
;    TITLE OF INVENTION:  A Rapid Means For Quantitating
;    TITLE OF INVENTION:  Genomic Instability
;    NUMBER OF SEQUENCES:  38
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hodgson, Russ, Andrews, Woods & Goodyear
;      STREET:  1800 One M&T Plaza
;      CITY:  Buffalo
;      STATE:  New York
;      COUNTRY:  United States
;      ZIP:  14203-2391
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS/ Microsoft Windows
;      SOFTWARE:  Wordperfect for Windows
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/734,973
;      FILING DATE:  October 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Nelson, M. Bud
;      REGISTRATION NUMBER:  35,300
;      REFERENCE/DOCKET NUMBER:  03551.0021
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (716) 856-4000
;      TELEFAX:  (716) 849-0349
;  INFORMATION FOR SEQ ID NO:  24 :
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10 nucleotides
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single-stranded
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
;    HYPOTHETICAL:  No
US-08-734-973-24

  Query Match             100.0%;  Score 8;  DB 2;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       1 CGCGCGCG 8


RESULT 32
US-08-734-973-24/c
; Sequence 24, Application US/08734973
; Patent No. 5912147
;  GENERAL INFORMATION:
;    APPLICANT:  Stoler, Daniel L.
;    APPLICANT:  Basik, Mark
;    APPLICANT:  Anderson, Garth R.
;    TITLE OF INVENTION:  A Rapid Means For Quantitating
;    TITLE OF INVENTION:  Genomic Instability
;    NUMBER OF SEQUENCES:  38
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hodgson, Russ, Andrews, Woods & Goodyear
;      STREET:  1800 One M&T Plaza
;      CITY:  Buffalo
;      STATE:  New York
;      COUNTRY:  United States
;      ZIP:  14203-2391
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS/ Microsoft Windows
;      SOFTWARE:  Wordperfect for Windows
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/734,973
;      FILING DATE:  October 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Nelson, M. Bud
;      REGISTRATION NUMBER:  35,300
;      REFERENCE/DOCKET NUMBER:  03551.0021
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (716) 856-4000
;      TELEFAX:  (716) 849-0349
;  INFORMATION FOR SEQ ID NO:  24 :
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10 nucleotides
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single-stranded
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
;    HYPOTHETICAL:  No
US-08-734-973-24

  Query Match             100.0%;  Score 8;  DB 2;  Length 10;
```

```
  Best Local Similarity   100.0%;  Pred. No. 1.2e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       8 CGCGCGCG 1


RESULT 33
US-08-734-973-25
; Sequence 25, Application US/08734973
; Patent No. 5912147
;  GENERAL INFORMATION:
;    APPLICANT:  Stoler, Daniel L.
;    APPLICANT:  Basik, Mark
;    APPLICANT:  Anderson, Garth R.
;    TITLE OF INVENTION:  A Rapid Means For Quantitating
;    TITLE OF INVENTION:  Genomic Instability
;    NUMBER OF SEQUENCES:  38
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hodgson, Russ, Andrews, Woods & Goodyear
;      STREET:  1800 One M&T Plaza
;      CITY:  Buffalo
;      STATE:  New York
;      COUNTRY:  United States
;      ZIP:  14203-2391
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS/ Microsoft Windows
;      SOFTWARE:  Wordperfect for Windows
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/734,973
;      FILING DATE:  October 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Nelson, M. Bud
;      REGISTRATION NUMBER:  35,300
;      REFERENCE/DOCKET NUMBER:  03551.0021
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (716) 856-4000
;      TELEFAX:  (716) 849-0349
;  INFORMATION FOR SEQ ID NO:  25 :
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10 nucleotides
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single-stranded
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
;    HYPOTHETICAL:  No
US-08-734-973-25

  Query Match             100.0%;  Score 8;  DB 2;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       1 CGCGCGCG 8


RESULT 34
US-08-734-973-25/c
; Sequence 25, Application US/08734973
; Patent No. 5912147
;  GENERAL INFORMATION:
;    APPLICANT:  Stoler, Daniel L.
;    APPLICANT:  Basik, Mark
;    APPLICANT:  Anderson, Garth R.
;    TITLE OF INVENTION:  A Rapid Means For Quantitating
;    TITLE OF INVENTION:  Genomic Instability
;    NUMBER OF SEQUENCES:  38
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hodgson, Russ, Andrews, Woods & Goodyear
;      STREET:  1800 One M&T Plaza
;      CITY:  Buffalo
;      STATE:  New York
;      COUNTRY:  United States
;      ZIP:  14203-2391
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS/ Microsoft Windows
;      SOFTWARE:  Wordperfect for Windows
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/734,973
;      FILING DATE:  October 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Nelson, M. Bud
;      REGISTRATION NUMBER:  35,300
;      REFERENCE/DOCKET NUMBER:  03551.0021
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (716) 856-4000
;      TELEFAX:  (716) 849-0349
;  INFORMATION FOR SEQ ID NO:  25 :
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10 nucleotides
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single-stranded
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
;    HYPOTHETICAL:  No
US-08-734-973-25

  Query Match             100.0%;  Score 8;  DB 2;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       8 CGCGCGCG 1


RESULT 35
US-08-734-973-26
; Sequence 26, Application US/08734973
; Patent No. 5912147
;  GENERAL INFORMATION:
;    APPLICANT:  Stoler, Daniel L.
;    APPLICANT:  Basik, Mark
;    APPLICANT:  Anderson, Garth R.
;    TITLE OF INVENTION:  A Rapid Means For Quantitating
;    TITLE OF INVENTION:  Genomic Instability
;    NUMBER OF SEQUENCES:  38
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hodgson, Russ, Andrews, Woods & Goodyear
;      STREET:  1800 One M&T Plaza
;      CITY:  Buffalo
;      STATE:  New York
;      COUNTRY:  United States
;      ZIP:  14203-2391
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS/ Microsoft Windows
;      SOFTWARE:  Wordperfect for Windows
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/734,973
;      FILING DATE:  October 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Nelson, M. Bud
;      REGISTRATION NUMBER:  35,300
;      REFERENCE/DOCKET NUMBER:  03551.0021
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (716) 856-4000
;      TELEFAX:  (716) 849-0349
;  INFORMATION FOR SEQ ID NO:  26 :
```

```
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10 nucleotides
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single-stranded
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
;    HYPOTHETICAL:  No
US-08-734-973-26

  Query Match             100.0%;  Score 8;  DB 2;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       1 CGCGCGCG 8


RESULT 36
US-08-734-973-26/c
; Sequence 26, Application US/08734973
; Patent No. 5912147
;  GENERAL INFORMATION:
;    APPLICANT:  Stoler, Daniel L.
;    APPLICANT:  Basik, Mark
;    APPLICANT:  Anderson, Garth R.
;    TITLE OF INVENTION:  A Rapid Means For Quantitating
;    TITLE OF INVENTION:  Genomic Instability
;    NUMBER OF SEQUENCES:  38
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hodgson, Russ, Andrews, Woods & Goodyear
;      STREET:  1800 One M&T Plaza
;      CITY:  Buffalo
;      STATE:  New York
;      COUNTRY:  United States
;      ZIP:  14203-2391
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS/ Microsoft Windows
;      SOFTWARE:  Wordperfect for Windows
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/734,973
;      FILING DATE:  October 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Nelson, M. Bud
;      REGISTRATION NUMBER:  35,300
;      REFERENCE/DOCKET NUMBER:  03551.0021
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (716) 856-4000
;      TELEFAX:  (716) 849-0349
;  INFORMATION FOR SEQ ID NO:  26 :
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10 nucleotides
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single-stranded
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
;    HYPOTHETICAL:  No
US-08-734-973-26

  Query Match             100.0%;  Score 8;  DB 2;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       8 CGCGCGCG 1


RESULT 37
US-08-734-973-27
; Sequence 27, Application US/08734973
; Patent No. 5912147
;  GENERAL INFORMATION:
;    APPLICANT:  Stoler, Daniel L.
;    APPLICANT:  Basik, Mark
;    APPLICANT:  Anderson, Garth R.
;    TITLE OF INVENTION:  A Rapid Means For Quantitating
;    TITLE OF INVENTION:  Genomic Instability
;    NUMBER OF SEQUENCES:  38
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hodgson, Russ, Andrews, Woods & Goodyear
;      STREET:  1800 One M&T Plaza
;      CITY:  Buffalo
;      STATE:  New York
;      COUNTRY:  United States
;      ZIP:  14203-2391
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS/ Microsoft Windows
;      SOFTWARE:  Wordperfect for Windows
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/734,973
;      FILING DATE:  October 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Nelson, M. Bud
;      REGISTRATION NUMBER:  35,300
;      REFERENCE/DOCKET NUMBER:  03551.0021
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (716) 856-4000
;      TELEFAX:  (716) 849-0349
;  INFORMATION FOR SEQ ID NO:  27 :
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10 nucleotides
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single-stranded
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
;    HYPOTHETICAL:  No
US-08-734-973-27

  Query Match             100.0%;  Score 8;  DB 2;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       1 CGCGCGCG 8


RESULT 38
US-08-734-973-27/c
; Sequence 27, Application US/08734973
; Patent No. 5912147
;  GENERAL INFORMATION:
;    APPLICANT:  Stoler, Daniel L.
;    APPLICANT:  Basik, Mark
;    APPLICANT:  Anderson, Garth R.
;    TITLE OF INVENTION:  A Rapid Means For Quantitating
;    TITLE OF INVENTION:  Genomic Instability
;    NUMBER OF SEQUENCES:  38
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hodgson, Russ, Andrews, Woods & Goodyear
;      STREET:  1800 One M&T Plaza
;      CITY:  Buffalo
;      STATE:  New York
;      COUNTRY:  United States
;      ZIP:  14203-2391
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS/ Microsoft Windows
;      SOFTWARE:  Wordperfect for Windows
;    CURRENT APPLICATION DATA:
```

```
;       APPLICATION NUMBER:  US/08/734,973
;       FILING DATE:  October 1996
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Nelson, M. Bud
;       REGISTRATION NUMBER:  35,300
;       REFERENCE/DOCKET NUMBER:  03551.0021
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (716) 856-4000
;       TELEFAX:  (716) 849-0349
;   INFORMATION FOR SEQ ID NO:  27 :
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  10 nucleotides
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single-stranded
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA
;     HYPOTHETICAL:  No
US-08-734-973-27

  Query Match             100.0%;  Score 8;  DB 2;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       8 CGCGCGCG 1


RESULT 39
US-08-729-598-13
; Sequence 13, Application US/08729598
; Patent No. 6001657
;  GENERAL INFORMATION:
;     APPLICANT:  Hardin, Charles C.
;     APPLICANT:  Brown II, Bernard A.
;     APPLICANT:  Roberts, John J.
;     APPLICANT:  Pelsue, Stephen A.
;     TITLE OF INVENTION:  Antibodies That Selectively Bind
;     TITLE OF INVENTION:  Quadruplex Nucleic Acids
;     NUMBER OF SEQUENCES:  13
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Sorojini J. Biswas
;       STREET:  P.O. Box 37428
;       CITY:  Raleigh
;       STATE:  No. 6001657th Carolina
;       COUNTRY:  USA
;       ZIP:  27627
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/729,598
;       FILING DATE:  11-OCT-1996
;       CLASSIFICATION:  530
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Biswas, Sorojini J.
;       REGISTRATION NUMBER:  39,111
;       REFERENCE/DOCKET NUMBER:  5051-301A
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (919) 854-1400
;       TELEFAX:  (919) 854-1401
;   INFORMATION FOR SEQ ID NO:  13:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  10 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  not relevant
;     MOLECULE TYPE:  DNA (genomic)
US-08-729-598-13

  Query Match             100.0%;  Score 8;  DB 3;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       1 CGCGCGCG 8


RESULT 40
US-08-729-598-13/c
; Sequence 13, Application US/08729598
; Patent No. 6001657
;  GENERAL INFORMATION:
;     APPLICANT:  Hardin, Charles C.
;     APPLICANT:  Brown II, Bernard A.
;     APPLICANT:  Roberts, John J.
;     APPLICANT:  Pelsue, Stephen A.
;     TITLE OF INVENTION:  Antibodies That Selectively Bind
;     TITLE OF INVENTION:  Quadruplex Nucleic Acids
;     NUMBER OF SEQUENCES:  13
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Sorojini J. Biswas
;       STREET:  P.O. Box 37428
;       CITY:  Raleigh
;       STATE:  No. 6001657th Carolina
;       COUNTRY:  USA
;       ZIP:  27627
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/729,598
;       FILING DATE:  11-OCT-1996
;       CLASSIFICATION:  530
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Biswas, Sorojini J.
;       REGISTRATION NUMBER:  39,111
;       REFERENCE/DOCKET NUMBER:  5051-301A
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (919) 854-1400
;       TELEFAX:  (919) 854-1401
;   INFORMATION FOR SEQ ID NO:  13:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  10 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  not relevant
;     MOLECULE TYPE:  DNA (genomic)
US-08-729-598-13

  Query Match             100.0%;  Score 8;  DB 3;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db      10 CGCGCGCG 3


RESULT 41
US-09-393-783A-79
; Sequence 79, Application US/09393783A
; Patent No. 6355428
; GENERAL INFORMATION:
;  APPLICANT: Schroth, Gary P.
;  APPLICANT:  Bruice, Thomas Wayne
;  APPLICANT:  Suh, Young J.
;  TITLE OF INVENTION: Nucleic Acid Ligand Interaction Assays
;  FILE REFERENCE: 4600-0128.30
;  CURRENT APPLICATION NUMBER: US/09/393,783A
;  CURRENT FILING DATE:  1999-10-09
```

```
;  PRIOR APPLICATION NUMBER: US 09/151,890
;  PRIOR FILING DATE: 1998-09-11
;  NUMBER OF SEQ ID NOS: 80
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 79
;   LENGTH: 12
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   NAME/KEY: misc_binding
;   LOCATION: (1)...(12)
;   OTHER INFORMATION: synthesized test oligonucleotide for binding
;   OTHER INFORMATION: studies
US-09-393-783A-79

  Query Match             100.0%;  Score 8;  DB 4;  Length 12;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       3 CGCGCGCG 10


RESULT 42
US-09-393-783A-79/c
; Sequence 79, Application US/09393783A
; Patent No. 6355428
; GENERAL INFORMATION:
;  APPLICANT: Schroth, Gary P.
;  APPLICANT:  Bruice, Thomas Wayne
;  APPLICANT:  Suh, Young J.
;  TITLE OF INVENTION: Nucleic Acid Ligand Interaction Assays
;  FILE REFERENCE: 4600-0128.30
;  CURRENT APPLICATION NUMBER: US/09/393,783A
;  CURRENT FILING DATE:  1999-10-09
;  PRIOR APPLICATION NUMBER: US 09/151,890
;  PRIOR FILING DATE: 1998-09-11
;  NUMBER OF SEQ ID NOS: 80
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 79
;   LENGTH: 12
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   NAME/KEY: misc_binding
;   LOCATION: (1)...(12)
;   OTHER INFORMATION: synthesized test oligonucleotide for binding
;   OTHER INFORMATION: studies
US-09-393-783A-79

  Query Match             100.0%;  Score 8;  DB 4;  Length 12;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db      10 CGCGCGCG 3


RESULT 43
US-09-151-890B-79
; Sequence 79, Application US/09151890B
; Patent No. 6420109
; GENERAL INFORMATION:
;  APPLICANT:  Gary P. Schroth
;  APPLICANT:  Thomas Wayne Bruice
;  APPLICANT:  Young J. Suh
;  TITLE OF INVENTION:   Nucleic Acid Ligand Interaction Assays
;  FILE REFERENCE:  4600-0128
;  CURRENT APPLICATION NUMBER: US/09/151,890B
;  CURRENT FILING DATE:  1998-09-11
;  NUMBER OF SEQ ID NOS:  80
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 79
;   LENGTH: 12
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   NAME/KEY: misc_binding
;   LOCATION: (1)...(12)
;   OTHER INFORMATION: synthesized test oligonucleotide for binding
;   OTHER INFORMATION: studies
US-09-151-890B-79

  Query Match             100.0%;  Score 8;  DB 4;  Length 12;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       3 CGCGCGCG 10


RESULT 44
US-09-151-890B-79/c
; Sequence 79, Application US/09151890B
; Patent No. 6420109
; GENERAL INFORMATION:
;  APPLICANT:  Gary P. Schroth
;  APPLICANT:  Thomas Wayne Bruice
;  APPLICANT:  Young J. Suh
;  TITLE OF INVENTION:   Nucleic Acid Ligand Interaction Assays
;  FILE REFERENCE:  4600-0128
;  CURRENT APPLICATION NUMBER: US/09/151,890B
;  CURRENT FILING DATE:  1998-09-11
;  NUMBER OF SEQ ID NOS:  80
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 79
;   LENGTH: 12
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   NAME/KEY: misc_binding
;   LOCATION: (1)...(12)
;   OTHER INFORMATION: synthesized test oligonucleotide for binding
;   OTHER INFORMATION: studies
US-09-151-890B-79

  Query Match             100.0%;  Score 8;  DB 4;  Length 12;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db      10 CGCGCGCG 3


RESULT 45
US-08-595-043A-44
; Sequence 44, Application US/08595043A
; Patent No. 5935824
;  GENERAL INFORMATION:
;    APPLICANT:  SGARLATO, GREGORY D.
;    TITLE OF INVENTION:  PROTEIN EXPRESSION SYSTEM
;    NUMBER OF SEQUENCES:  90
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
```

```
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/595,043A
;       FILING DATE:  31-JAN-1996
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  CARROLL, PETER G.
;       REGISTRATION NUMBER:  32,837
;       REFERENCE/DOCKET NUMBER:  SGAR-00371
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (415) 705-8410
;       TELEFAX:  (415) 397-8338
;   INFORMATION FOR SEQ ID NO:  44:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  14 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
US-08-595-043A-44

  Query Match             100.0%;  Score 8;  DB 2;  Length 14;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       6 CGCGCGCG 13


Search completed: March 14, 2003, 04:33:32
Job time : 44 secs
```

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: March 14, 2003, 02:36:13 ; Search time 1363 Seconds
(without alignments)
170.816 Million cell updates/sec

Title: CGCGCGCG
Perfect score: 8
Sequence: 1 cgcgcgcg 8

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	8	100.0	10	6	AR071780	AR071780 Sequence
c	2	8	100.0	10	6	AR071780	AR071780 Sequence
	3	8	100.0	10	6	AR071781	AR071781 Sequence
c	4	8	100.0	10	6	AR071781	AR071781 Sequence
	5	8	100.0	10	6	AR071782	AR071782 Sequence
c	6	8	100.0	10	6	AR071782	AR071782 Sequence
	7	8	100.0	10	6	AR071783	AR071783 Sequence
c	8	8	100.0	10	6	AR071783	AR071783 Sequence
	9	8	100.0	10	6	AR071784	AR071784 Sequence
c	10	8	100.0	10	6	AR071784	AR071784 Sequence
	11	8	100.0	10	6	AR071785	AR071785 Sequence
c	12	8	100.0	10	6	AR071785	AR071785 Sequence
	13	8	100.0	10	6	AR071786	AR071786 Sequence
c	14	8	100.0	10	6	AR071786	AR071786 Sequence
	15	8	100.0	10	6	AR071787	AR071787 Sequence
c	16	8	100.0	10	6	AR071787	AR071787 Sequence
	17	8	100.0	10	6	AR071788	AR071788 Sequence
c	18	8	100.0	10	6	AR071788	AR071788 Sequence
	19	8	100.0	10	6	AR071789	AR071789 Sequence
c	20	8	100.0	10	6	AR071789	AR071789 Sequence
	21	8	100.0	10	6	AR071790	AR071790 Sequence
c	22	8	100.0	10	6	AR071790	AR071790 Sequence
	23	8	100.0	10	6	AR071791	AR071791 Sequence
c	24	8	100.0	10	6	AR071791	AR071791 Sequence
	25	8	100.0	10	6	AR071792	AR071792 Sequence
c	26	8	100.0	10	6	AR071792	AR071792 Sequence
	27	8	100.0	10	6	AR071793	AR071793 Sequence
c	28	8	100.0	10	6	AR071793	AR071793 Sequence
	29	8	100.0	10	6	AR071794	AR071794 Sequence
c	30	8	100.0	10	6	AR071794	AR071794 Sequence
	31	8	100.0	10	6	AR071795	AR071795 Sequence
c	32	8	100.0	10	6	AR071795	AR071795 Sequence
	33	8	100.0	10	6	AR071796	AR071796 Sequence
c	34	8	100.0	10	6	AR071796	AR071796 Sequence
	35	8	100.0	10	6	AR071797	AR071797 Sequence
c	36	8	100.0	10	6	AR071797	AR071797 Sequence
	37	8	100.0	10	6	AR071798	AR071798 Sequence
c	38	8	100.0	10	6	AR071798	AR071798 Sequence
	39	8	100.0	10	6	AR094565	AR094565 Sequence
c	40	8	100.0	10	6	AR094565	AR094565 Sequence
	41	8	100.0	12	6	AR199370	AR199370 Sequence
c	42	8	100.0	12	6	AR199370	AR199370 Sequence
	43	8	100.0	16	6	AR048207	AR048207 Sequence
c	44	8	100.0	16	6	AR048207	AR048207 Sequence
	45	8	100.0	16	6	AR055680	AR055680 Sequence

ALIGNMENTS

RESULT 1
AR071780
LOCUS AR071780 10 bp DNA linear PAT 18-FEB-2000
DEFINITION Sequence 9 from patent US 5912147.
ACCESSION AR071780
VERSION AR071780.1 GI:7222668
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 10)
AUTHORS Stoler,D., Basik,M. and Anderson,G.
TITLE Rapid means of quantitating genomic instability
JOURNAL Patent: US 5912147-A 9 15-JUN-1999;
FEATURES Location/Qualifiers

```
     source          1. .10
                     /organism="unknown"
BASE COUNT        1 a      4 c      5 g      0 t
ORIGIN

  Query Match             100.0%;  Score 8;  DB 6;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 8.9e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db      1 CGCGCGCG 8


RESULT 2
AR071780/c
LOCUS       AR071780                  10 bp    DNA     linear   PAT 18-FEB-2000
DEFINITION  Sequence 9 from patent US 5912147.
ACCESSION   AR071780
VERSION     AR071780.1  GI:7222668
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 10)
  AUTHORS   Stoler,D., Basik,M. and Anderson,G.
  TITLE     Rapid means of quantitating genomic instability
  JOURNAL   Patent: US 5912147-A 9 15-JUN-1999;
FEATURES             Location/Qualifiers
     source          1. .10
                     /organism="unknown"
BASE COUNT        1 a      4 c      5 g      0 t
ORIGIN

  Query Match             100.0%;  Score 8;  DB 6;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 8.9e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db      8 CGCGCGCG 1


RESULT 3
AR071781
LOCUS       AR071781                  10 bp    DNA     linear   PAT 18-FEB-2000
DEFINITION  Sequence 10 from patent US 5912147.
ACCESSION   AR071781
VERSION     AR071781.1  GI:7222669
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 10)
  AUTHORS   Stoler,D., Basik,M. and Anderson,G.
  TITLE     Rapid means of quantitating genomic instability
  JOURNAL   Patent: US 5912147-A 10 15-JUN-1999;
FEATURES             Location/Qualifiers
     source          1. .10
                     /organism="unknown"
BASE COUNT        0 a      4 c      6 g      0 t
ORIGIN

  Query Match             100.0%;  Score 8;  DB 6;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 8.9e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db      1 CGCGCGCG 8


RESULT 4
```

```
AR071781/c
LOCUS       AR071781                  10 bp    DNA     linear   PAT 18-FEB-2000
DEFINITION  Sequence 10 from patent US 5912147.
ACCESSION   AR071781
VERSION     AR071781.1  GI:7222669
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 10)
  AUTHORS   Stoler,D., Basik,M. and Anderson,G.
  TITLE     Rapid means of quantitating genomic instability
  JOURNAL   Patent: US 5912147-A 10 15-JUN-1999;
FEATURES             Location/Qualifiers
     source          1. .10
                     /organism="unknown"
BASE COUNT        0 a      4 c      6 g      0 t
ORIGIN

  Query Match             100.0%;  Score 8;  DB 6;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 8.9e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db      8 CGCGCGCG 1


RESULT 5
AR071782
LOCUS       AR071782                  10 bp    DNA     linear   PAT 18-FEB-2000
DEFINITION  Sequence 11 from patent US 5912147.
ACCESSION   AR071782
VERSION     AR071782.1  GI:7222670
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 10)
  AUTHORS   Stoler,D., Basik,M. and Anderson,G.
  TITLE     Rapid means of quantitating genomic instability
  JOURNAL   Patent: US 5912147-A 11 15-JUN-1999;
FEATURES             Location/Qualifiers
     source          1. .10
                     /organism="unknown"
BASE COUNT        1 a      5 c      4 g      0 t
ORIGIN

  Query Match             100.0%;  Score 8;  DB 6;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 8.9e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db      1 CGCGCGCG 8


RESULT 6
AR071782/c
LOCUS       AR071782                  10 bp    DNA     linear   PAT 18-FEB-2000
DEFINITION  Sequence 11 from patent US 5912147.
ACCESSION   AR071782
VERSION     AR071782.1  GI:7222670
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 10)
  AUTHORS   Stoler,D., Basik,M. and Anderson,G.
  TITLE     Rapid means of quantitating genomic instability
  JOURNAL   Patent: US 5912147-A 11 15-JUN-1999;
FEATURES             Location/Qualifiers
     source          1. .10
```

```
                     /organism="unknown"
BASE COUNT        1 a      5 c      4 g      0 t
ORIGIN

  Query Match            100.0%;  Score 8;  DB 6;  Length 10;
  Best Local Similarity  100.0%;  Pred. No. 8.9e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db      8 CGCGCGCG 1


RESULT 7
AR071783
LOCUS       AR071783                 10 bp    DNA     linear   PAT 18-FEB-2000
DEFINITION  Sequence 12 from patent US 5912147.
ACCESSION   AR071783
VERSION     AR071783.1  GI:7222671
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 10)
  AUTHORS   Stoler,D., Basik,M. and Anderson,G.
  TITLE     Rapid means of quantitating genomic instability
  JOURNAL   Patent: US 5912147-A 12 15-JUN-1999;
FEATURES             Location/Qualifiers
     source          1. .10
                     /organism="unknown"
BASE COUNT        1 a      4 c      4 g      1 t
ORIGIN

  Query Match            100.0%;  Score 8;  DB 6;  Length 10;
  Best Local Similarity  100.0%;  Pred. No. 8.9e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db      1 CGCGCGCG 8


RESULT 8
AR071783/c
LOCUS       AR071783                 10 bp    DNA     linear   PAT 18-FEB-2000
DEFINITION  Sequence 12 from patent US 5912147.
ACCESSION   AR071783
VERSION     AR071783.1  GI:7222671
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 10)
  AUTHORS   Stoler,D., Basik,M. and Anderson,G.
  TITLE     Rapid means of quantitating genomic instability
  JOURNAL   Patent: US 5912147-A 12 15-JUN-1999;
FEATURES             Location/Qualifiers
     source          1. .10
                     /organism="unknown"
BASE COUNT        1 a      4 c      4 g      1 t
ORIGIN

  Query Match            100.0%;  Score 8;  DB 6;  Length 10;
  Best Local Similarity  100.0%;  Pred. No. 8.9e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db      8 CGCGCGCG 1


RESULT 9
AR071784
LOCUS       AR071784                 10 bp    DNA     linear   PAT 18-FEB-2000
DEFINITION  Sequence 13 from patent US 5912147.
ACCESSION   AR071784
VERSION     AR071784.1  GI:7222672
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 10)
  AUTHORS   Stoler,D., Basik,M. and Anderson,G.
  TITLE     Rapid means of quantitating genomic instability
  JOURNAL   Patent: US 5912147-A 13 15-JUN-1999;
FEATURES             Location/Qualifiers
     source          1. .10
                     /organism="unknown"
BASE COUNT        1 a      4 c      4 g      1 t
ORIGIN

  Query Match            100.0%;  Score 8;  DB 6;  Length 10;
  Best Local Similarity  100.0%;  Pred. No. 8.9e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db      1 CGCGCGCG 8


RESULT 10
AR071784/c
LOCUS       AR071784                 10 bp    DNA     linear   PAT 18-FEB-2000
DEFINITION  Sequence 13 from patent US 5912147.
ACCESSION   AR071784
VERSION     AR071784.1  GI:7222672
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 10)
  AUTHORS   Stoler,D., Basik,M. and Anderson,G.
  TITLE     Rapid means of quantitating genomic instability
  JOURNAL   Patent: US 5912147-A 13 15-JUN-1999;
FEATURES             Location/Qualifiers
     source          1. .10
                     /organism="unknown"
BASE COUNT        1 a      4 c      4 g      1 t
ORIGIN

  Query Match            100.0%;  Score 8;  DB 6;  Length 10;
  Best Local Similarity  100.0%;  Pred. No. 8.9e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db      8 CGCGCGCG 1


RESULT 11
AR071785
LOCUS       AR071785                 10 bp    DNA     linear   PAT 18-FEB-2000
DEFINITION  Sequence 14 from patent US 5912147.
ACCESSION   AR071785
VERSION     AR071785.1  GI:7222673
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 10)
  AUTHORS   Stoler,D., Basik,M. and Anderson,G.
  TITLE     Rapid means of quantitating genomic instability
  JOURNAL   Patent: US 5912147-A 14 15-JUN-1999;
FEATURES             Location/Qualifiers
     source          1. .10
                     /organism="unknown"
```

```
BASE COUNT        0 a      5 c      5 g      0 t
ORIGIN

  Query Match             100.0%;  Score 8;  DB 6;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 8.9e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db      1 CGCGCGCG 8


RESULT 12
AR071785/c
LOCUS       AR071785                 10 bp    DNA     linear   PAT 18-FEB-2000
DEFINITION  Sequence 14 from patent US 5912147.
ACCESSION   AR071785
VERSION     AR071785.1  GI:7222673
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 10)
  AUTHORS   Stoler,D., Basik,M. and Anderson,G.
  TITLE     Rapid means of quantitating genomic instability
  JOURNAL   Patent: US 5912147-A 14 15-JUN-1999;
FEATURES             Location/Qualifiers
     source          1. .10
                     /organism="unknown"
BASE COUNT        0 a      5 c      5 g      0 t
ORIGIN

  Query Match             100.0%;  Score 8;  DB 6;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 8.9e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db      8 CGCGCGCG 1


RESULT 13
AR071786
LOCUS       AR071786                 10 bp    DNA     linear   PAT 18-FEB-2000
DEFINITION  Sequence 15 from patent US 5912147.
ACCESSION   AR071786
VERSION     AR071786.1  GI:7222674
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 10)
  AUTHORS   Stoler,D., Basik,M. and Anderson,G.
  TITLE     Rapid means of quantitating genomic instability
  JOURNAL   Patent: US 5912147-A 15 15-JUN-1999;
FEATURES             Location/Qualifiers
     source          1. .10
                     /organism="unknown"
BASE COUNT        0 a      4 c      5 g      1 t
ORIGIN

  Query Match             100.0%;  Score 8;  DB 6;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 8.9e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db      1 CGCGCGCG 8


RESULT 14
AR071786/c
LOCUS       AR071786                 10 bp    DNA     linear   PAT 18-FEB-2000
DEFINITION  Sequence 15 from patent US 5912147.
ACCESSION   AR071786
VERSION     AR071786.1  GI:7222674
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 10)
  AUTHORS   Stoler,D., Basik,M. and Anderson,G.
  TITLE     Rapid means of quantitating genomic instability
  JOURNAL   Patent: US 5912147-A 15 15-JUN-1999;
FEATURES             Location/Qualifiers
     source          1. .10
                     /organism="unknown"
BASE COUNT        0 a      4 c      5 g      1 t
ORIGIN

  Query Match             100.0%;  Score 8;  DB 6;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 8.9e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db      8 CGCGCGCG 1


RESULT 15
AR071787
LOCUS       AR071787                 10 bp    DNA     linear   PAT 18-FEB-2000
DEFINITION  Sequence 16 from patent US 5912147.
ACCESSION   AR071787
VERSION     AR071787.1  GI:7222675
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 10)
  AUTHORS   Stoler,D., Basik,M. and Anderson,G.
  TITLE     Rapid means of quantitating genomic instability
  JOURNAL   Patent: US 5912147-A 16 15-JUN-1999;
FEATURES             Location/Qualifiers
     source          1. .10
                     /organism="unknown"
BASE COUNT        0 a      4 c      5 g      1 t
ORIGIN

  Query Match             100.0%;  Score 8;  DB 6;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 8.9e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db      1 CGCGCGCG 8


RESULT 16
AR071787/c
LOCUS       AR071787                 10 bp    DNA     linear   PAT 18-FEB-2000
DEFINITION  Sequence 16 from patent US 5912147.
ACCESSION   AR071787
VERSION     AR071787.1  GI:7222675
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 10)
  AUTHORS   Stoler,D., Basik,M. and Anderson,G.
  TITLE     Rapid means of quantitating genomic instability
  JOURNAL   Patent: US 5912147-A 16 15-JUN-1999;
FEATURES             Location/Qualifiers
     source          1. .10
                     /organism="unknown"
BASE COUNT        0 a      4 c      5 g      1 t
```

```
ORIGIN

  Query Match             100.0%;  Score 8;  DB 6;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 8.9e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db      8 CGCGCGCG 1


RESULT 17
AR071788
LOCUS       AR071788                  10 bp    DNA     linear   PAT 18-FEB-2000
DEFINITION  Sequence 17 from patent US 5912147.
ACCESSION   AR071788
VERSION     AR071788.1  GI:7222676
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 10)
  AUTHORS   Stoler,D., Basik,M. and Anderson,G.
  TITLE     Rapid means of quantitating genomic instability
  JOURNAL   Patent: US 5912147-A 17 15-JUN-1999;
FEATURES             Location/Qualifiers
     source          1. .10
                     /organism="unknown"
BASE COUNT        2 a      4 c      4 g      0 t
ORIGIN

  Query Match             100.0%;  Score 8;  DB 6;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 8.9e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db      1 CGCGCGCG 8


RESULT 18
AR071788/c
LOCUS       AR071788                  10 bp    DNA     linear   PAT 18-FEB-2000
DEFINITION  Sequence 17 from patent US 5912147.
ACCESSION   AR071788
VERSION     AR071788.1  GI:7222676
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 10)
  AUTHORS   Stoler,D., Basik,M. and Anderson,G.
  TITLE     Rapid means of quantitating genomic instability
  JOURNAL   Patent: US 5912147-A 17 15-JUN-1999;
FEATURES             Location/Qualifiers
     source          1. .10
                     /organism="unknown"
BASE COUNT        2 a      4 c      4 g      0 t
ORIGIN

  Query Match             100.0%;  Score 8;  DB 6;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 8.9e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db      8 CGCGCGCG 1


RESULT 19
AR071789
LOCUS       AR071789                  10 bp    DNA     linear   PAT 18-FEB-2000
DEFINITION  Sequence 18 from patent US 5912147.
ACCESSION   AR071789
VERSION     AR071789.1  GI:7222677
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 10)
  AUTHORS   Stoler,D., Basik,M. and Anderson,G.
  TITLE     Rapid means of quantitating genomic instability
  JOURNAL   Patent: US 5912147-A 18 15-JUN-1999;
FEATURES             Location/Qualifiers
     source          1. .10
                     /organism="unknown"
BASE COUNT        1 a      4 c      5 g      0 t
ORIGIN

  Query Match             100.0%;  Score 8;  DB 6;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 8.9e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db      1 CGCGCGCG 8


RESULT 20
AR071789/c
LOCUS       AR071789                  10 bp    DNA     linear   PAT 18-FEB-2000
DEFINITION  Sequence 18 from patent US 5912147.
ACCESSION   AR071789
VERSION     AR071789.1  GI:7222677
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 10)
  AUTHORS   Stoler,D., Basik,M. and Anderson,G.
  TITLE     Rapid means of quantitating genomic instability
  JOURNAL   Patent: US 5912147-A 18 15-JUN-1999;
FEATURES             Location/Qualifiers
     source          1. .10
                     /organism="unknown"
BASE COUNT        1 a      4 c      5 g      0 t
ORIGIN

  Query Match             100.0%;  Score 8;  DB 6;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 8.9e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db      8 CGCGCGCG 1


RESULT 21
AR071790
LOCUS       AR071790                  10 bp    DNA     linear   PAT 18-FEB-2000
DEFINITION  Sequence 19 from patent US 5912147.
ACCESSION   AR071790
VERSION     AR071790.1  GI:7222678
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 10)
  AUTHORS   Stoler,D., Basik,M. and Anderson,G.
  TITLE     Rapid means of quantitating genomic instability
  JOURNAL   Patent: US 5912147-A 19 15-JUN-1999;
FEATURES             Location/Qualifiers
     source          1. .10
                     /organism="unknown"
BASE COUNT        0 a      6 c      4 g      0 t
ORIGIN
```

```
  Query Match             100.0%;  Score 8;  DB 6;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 8.9e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db      1 CGCGCGCG 8


RESULT 22
AR071790/c
LOCUS       AR071790                  10 bp    DNA     linear   PAT 18-FEB-2000
DEFINITION  Sequence 19 from patent US 5912147.
ACCESSION   AR071790
VERSION     AR071790.1  GI:7222678
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 10)
  AUTHORS   Stoler,D., Basik,M. and Anderson,G.
  TITLE     Rapid means of quantitating genomic instability
  JOURNAL   Patent: US 5912147-A 19 15-JUN-1999;
FEATURES             Location/Qualifiers
     source          1. .10
                     /organism="unknown"
BASE COUNT        0 a      6 c      4 g      0 t
ORIGIN

  Query Match             100.0%;  Score 8;  DB 6;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 8.9e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db      8 CGCGCGCG 1


RESULT 23
AR071791
LOCUS       AR071791                  10 bp    DNA     linear   PAT 18-FEB-2000
DEFINITION  Sequence 20 from patent US 5912147.
ACCESSION   AR071791
VERSION     AR071791.1  GI:7222679
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 10)
  AUTHORS   Stoler,D., Basik,M. and Anderson,G.
  TITLE     Rapid means of quantitating genomic instability
  JOURNAL   Patent: US 5912147-A 20 15-JUN-1999;
FEATURES             Location/Qualifiers
     source          1. .10
                     /organism="unknown"
BASE COUNT        0 a      5 c      4 g      1 t
ORIGIN

  Query Match             100.0%;  Score 8;  DB 6;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 8.9e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db      1 CGCGCGCG 8


RESULT 24
AR071791/c
LOCUS       AR071791                  10 bp    DNA     linear   PAT 18-FEB-2000
DEFINITION  Sequence 20 from patent US 5912147.
ACCESSION   AR071791
VERSION     AR071791.1  GI:7222679
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 10)
  AUTHORS   Stoler,D., Basik,M. and Anderson,G.
  TITLE     Rapid means of quantitating genomic instability
  JOURNAL   Patent: US 5912147-A 20 15-JUN-1999;
FEATURES             Location/Qualifiers
     source          1. .10
                     /organism="unknown"
BASE COUNT        0 a      5 c      4 g      1 t
ORIGIN

  Query Match             100.0%;  Score 8;  DB 6;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 8.9e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db      8 CGCGCGCG 1


RESULT 25
AR071792
LOCUS       AR071792                  10 bp    DNA     linear   PAT 18-FEB-2000
DEFINITION  Sequence 21 from patent US 5912147.
ACCESSION   AR071792
VERSION     AR071792.1  GI:7222680
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 10)
  AUTHORS   Stoler,D., Basik,M. and Anderson,G.
  TITLE     Rapid means of quantitating genomic instability
  JOURNAL   Patent: US 5912147-A 21 15-JUN-1999;
FEATURES             Location/Qualifiers
     source          1. .10
                     /organism="unknown"
BASE COUNT        0 a      5 c      4 g      1 t
ORIGIN

  Query Match             100.0%;  Score 8;  DB 6;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 8.9e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db      1 CGCGCGCG 8


RESULT 26
AR071792/c
LOCUS       AR071792                  10 bp    DNA     linear   PAT 18-FEB-2000
DEFINITION  Sequence 21 from patent US 5912147.
ACCESSION   AR071792
VERSION     AR071792.1  GI:7222680
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 10)
  AUTHORS   Stoler,D., Basik,M. and Anderson,G.
  TITLE     Rapid means of quantitating genomic instability
  JOURNAL   Patent: US 5912147-A 21 15-JUN-1999;
FEATURES             Location/Qualifiers
     source          1. .10
                     /organism="unknown"
BASE COUNT        0 a      5 c      4 g      1 t
ORIGIN
```

```
  Query Match            100.0%;  Score 8;  DB 6;  Length 10;
  Best Local Similarity  100.0%;  Pred. No. 8.9e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db      8 CGCGCGCG 1


RESULT 27
AR071793
LOCUS       AR071793                 10 bp    DNA     linear   PAT 18-FEB-2000
DEFINITION  Sequence 22 from patent US 5912147.
ACCESSION   AR071793
VERSION     AR071793.1  GI:7222681
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 10)
  AUTHORS   Stoler,D., Basik,M. and Anderson,G.
  TITLE     Rapid means of quantitating genomic instability
  JOURNAL   Patent: US 5912147-A 22 15-JUN-1999;
FEATURES             Location/Qualifiers
     source          1. .10
                     /organism="unknown"
BASE COUNT        0 a      4 c      4 g      2 t
ORIGIN

  Query Match            100.0%;  Score 8;  DB 6;  Length 10;
  Best Local Similarity  100.0%;  Pred. No. 8.9e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db      1 CGCGCGCG 8


RESULT 28
AR071793/c
LOCUS       AR071793                 10 bp    DNA     linear   PAT 18-FEB-2000
DEFINITION  Sequence 22 from patent US 5912147.
ACCESSION   AR071793
VERSION     AR071793.1  GI:7222681
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 10)
  AUTHORS   Stoler,D., Basik,M. and Anderson,G.
  TITLE     Rapid means of quantitating genomic instability
  JOURNAL   Patent: US 5912147-A 22 15-JUN-1999;
FEATURES             Location/Qualifiers
     source          1. .10
                     /organism="unknown"
BASE COUNT        0 a      4 c      4 g      2 t
ORIGIN

  Query Match            100.0%;  Score 8;  DB 6;  Length 10;
  Best Local Similarity  100.0%;  Pred. No. 8.9e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db      8 CGCGCGCG 1


RESULT 29
AR071794
LOCUS       AR071794                 10 bp    DNA     linear   PAT 18-FEB-2000
DEFINITION  Sequence 23 from patent US 5912147.
ACCESSION   AR071794
VERSION     AR071794.1  GI:7222682
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 10)
  AUTHORS   Stoler,D., Basik,M. and Anderson,G.
  TITLE     Rapid means of quantitating genomic instability
  JOURNAL   Patent: US 5912147-A 23 15-JUN-1999;
FEATURES             Location/Qualifiers
     source          1. .10
                     /organism="unknown"
BASE COUNT        0 a      5 c      4 g      1 t
ORIGIN

  Query Match            100.0%;  Score 8;  DB 6;  Length 10;
  Best Local Similarity  100.0%;  Pred. No. 8.9e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db      1 CGCGCGCG 8


RESULT 30
AR071794/c
LOCUS       AR071794                 10 bp    DNA     linear   PAT 18-FEB-2000
DEFINITION  Sequence 23 from patent US 5912147.
ACCESSION   AR071794
VERSION     AR071794.1  GI:7222682
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 10)
  AUTHORS   Stoler,D., Basik,M. and Anderson,G.
  TITLE     Rapid means of quantitating genomic instability
  JOURNAL   Patent: US 5912147-A 23 15-JUN-1999;
FEATURES             Location/Qualifiers
     source          1. .10
                     /organism="unknown"
BASE COUNT        0 a      5 c      4 g      1 t
ORIGIN

  Query Match            100.0%;  Score 8;  DB 6;  Length 10;
  Best Local Similarity  100.0%;  Pred. No. 8.9e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db      8 CGCGCGCG 1


RESULT 31
AR071795
LOCUS       AR071795                 10 bp    DNA     linear   PAT 18-FEB-2000
DEFINITION  Sequence 24 from patent US 5912147.
ACCESSION   AR071795
VERSION     AR071795.1  GI:7222683
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 10)
  AUTHORS   Stoler,D., Basik,M. and Anderson,G.
  TITLE     Rapid means of quantitating genomic instability
  JOURNAL   Patent: US 5912147-A 24 15-JUN-1999;
FEATURES             Location/Qualifiers
     source          1. .10
                     /organism="unknown"
BASE COUNT        0 a      4 c      4 g      2 t
ORIGIN

  Query Match            100.0%;  Score 8;  DB 6;  Length 10;
```

```
  Best Local Similarity   100.0%;  Pred. No. 8.9e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db      1 CGCGCGCG 8


RESULT 32
AR071795/c
LOCUS       AR071795                 10 bp    DNA     linear   PAT 18-FEB-2000
DEFINITION  Sequence 24 from patent US 5912147.
ACCESSION   AR071795
VERSION     AR071795.1  GI:7222683
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 10)
  AUTHORS   Stoler,D., Basik,M. and Anderson,G.
  TITLE     Rapid means of quantitating genomic instability
  JOURNAL   Patent: US 5912147-A 24 15-JUN-1999;
FEATURES             Location/Qualifiers
     source          1. .10
                     /organism="unknown"
BASE COUNT        0 a      4 c      4 g      2 t
ORIGIN

  Query Match             100.0%;  Score 8;  DB 6;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 8.9e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db      8 CGCGCGCG 1


RESULT 33
AR071796
LOCUS       AR071796                 10 bp    DNA     linear   PAT 18-FEB-2000
DEFINITION  Sequence 25 from patent US 5912147.
ACCESSION   AR071796
VERSION     AR071796.1  GI:7222684
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 10)
  AUTHORS   Stoler,D., Basik,M. and Anderson,G.
  TITLE     Rapid means of quantitating genomic instability
  JOURNAL   Patent: US 5912147-A 25 15-JUN-1999;
FEATURES             Location/Qualifiers
     source          1. .10
                     /organism="unknown"
BASE COUNT        0 a      5 c      4 g      1 t
ORIGIN

  Query Match             100.0%;  Score 8;  DB 6;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 8.9e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db      1 CGCGCGCG 8


RESULT 34
AR071796/c
LOCUS       AR071796                 10 bp    DNA     linear   PAT 18-FEB-2000
DEFINITION  Sequence 25 from patent US 5912147.
ACCESSION   AR071796
VERSION     AR071796.1  GI:7222684
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 10)
  AUTHORS   Stoler,D., Basik,M. and Anderson,G.
  TITLE     Rapid means of quantitating genomic instability
  JOURNAL   Patent: US 5912147-A 25 15-JUN-1999;
FEATURES             Location/Qualifiers
     source          1. .10
                     /organism="unknown"
BASE COUNT        0 a      5 c      4 g      1 t
ORIGIN

  Query Match             100.0%;  Score 8;  DB 6;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 8.9e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db      8 CGCGCGCG 1


RESULT 35
AR071797
LOCUS       AR071797                 10 bp    DNA     linear   PAT 18-FEB-2000
DEFINITION  Sequence 26 from patent US 5912147.
ACCESSION   AR071797
VERSION     AR071797.1  GI:7222685
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 10)
  AUTHORS   Stoler,D., Basik,M. and Anderson,G.
  TITLE     Rapid means of quantitating genomic instability
  JOURNAL   Patent: US 5912147-A 26 15-JUN-1999;
FEATURES             Location/Qualifiers
     source          1. .10
                     /organism="unknown"
BASE COUNT        0 a      4 c      4 g      2 t
ORIGIN

  Query Match             100.0%;  Score 8;  DB 6;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 8.9e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db      1 CGCGCGCG 8


RESULT 36
AR071797/c
LOCUS       AR071797                 10 bp    DNA     linear   PAT 18-FEB-2000
DEFINITION  Sequence 26 from patent US 5912147.
ACCESSION   AR071797
VERSION     AR071797.1  GI:7222685
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 10)
  AUTHORS   Stoler,D., Basik,M. and Anderson,G.
  TITLE     Rapid means of quantitating genomic instability
  JOURNAL   Patent: US 5912147-A 26 15-JUN-1999;
FEATURES             Location/Qualifiers
     source          1. .10
                     /organism="unknown"
BASE COUNT        0 a      4 c      4 g      2 t
ORIGIN

  Query Match             100.0%;  Score 8;  DB 6;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 8.9e+05;
```

```
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db      8 CGCGCGCG 1


RESULT 37
AR071798
LOCUS       AR071798                 10 bp    DNA     linear   PAT 18-FEB-2000
DEFINITION  Sequence 27 from patent US 5912147.
ACCESSION   AR071798
VERSION     AR071798.1  GI:7222686
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 10)
  AUTHORS   Stoler,D., Basik,M. and Anderson,G.
  TITLE     Rapid means of quantitating genomic instability
  JOURNAL   Patent: US 5912147-A 27 15-JUN-1999;
FEATURES             Location/Qualifiers
     source          1. .10
                     /organism="unknown"
BASE COUNT        0 a      4 c      4 g      2 t
ORIGIN

  Query Match           100.0%;  Score 8;  DB 6;  Length 10;
  Best Local Similarity  100.0%;  Pred. No. 8.9e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db      1 CGCGCGCG 8


RESULT 38
AR071798/c
LOCUS       AR071798                 10 bp    DNA     linear   PAT 18-FEB-2000
DEFINITION  Sequence 27 from patent US 5912147.
ACCESSION   AR071798
VERSION     AR071798.1  GI:7222686
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 10)
  AUTHORS   Stoler,D., Basik,M. and Anderson,G.
  TITLE     Rapid means of quantitating genomic instability
  JOURNAL   Patent: US 5912147-A 27 15-JUN-1999;
FEATURES             Location/Qualifiers
     source          1. .10
                     /organism="unknown"
BASE COUNT        0 a      4 c      4 g      2 t
ORIGIN

  Query Match           100.0%;  Score 8;  DB 6;  Length 10;
  Best Local Similarity  100.0%;  Pred. No. 8.9e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db      8 CGCGCGCG 1


RESULT 39
AR094565
LOCUS       AR094565                 10 bp    DNA     linear   PAT 08-SEP-2000
DEFINITION  Sequence 13 from patent US 6001657.
ACCESSION   AR094565
VERSION     AR094565.1  GI:10021597
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 10)
  AUTHORS   Hardin,C.C., Brown,B.A. II, Roberts,J.F. and Pelsue,S.C.
  TITLE     Antibodies that selectively bind quadruplex nucleic acids
  JOURNAL   Patent: US 6001657-A 13 14-DEC-1999;
FEATURES             Location/Qualifiers
     source          1. .10
                     /organism="unknown"
BASE COUNT        0 a      5 c      5 g      0 t
ORIGIN

  Query Match           100.0%;  Score 8;  DB 6;  Length 10;
  Best Local Similarity  100.0%;  Pred. No. 8.9e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db      1 CGCGCGCG 8


RESULT 40
AR094565/c
LOCUS       AR094565                 10 bp    DNA     linear   PAT 08-SEP-2000
DEFINITION  Sequence 13 from patent US 6001657.
ACCESSION   AR094565
VERSION     AR094565.1  GI:10021597
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 10)
  AUTHORS   Hardin,C.C., Brown,B.A. II, Roberts,J.F. and Pelsue,S.C.
  TITLE     Antibodies that selectively bind quadruplex nucleic acids
  JOURNAL   Patent: US 6001657-A 13 14-DEC-1999;
FEATURES             Location/Qualifiers
     source          1. .10
                     /organism="unknown"
BASE COUNT        0 a      5 c      5 g      0 t
ORIGIN

  Query Match           100.0%;  Score 8;  DB 6;  Length 10;
  Best Local Similarity  100.0%;  Pred. No. 8.9e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db     10 CGCGCGCG 3


RESULT 41
AR199370
LOCUS       AR199370                 12 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 79 from patent US 6355428.
ACCESSION   AR199370
VERSION     AR199370.1  GI:20249444
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 12)
  AUTHORS   Schroth,G.P., Bruice,T.Wayne. and Suh,Y.J.
  TITLE     Nucleic acid ligand interaction assays
  JOURNAL   Patent: US 6355428-A 79 12-MAR-2002;
FEATURES             Location/Qualifiers
     source          1. .12
                     /organism="unknown"
BASE COUNT        0 a      8 c      4 g      0 t
ORIGIN

  Query Match           100.0%;  Score 8;  DB 6;  Length 12;
  Best Local Similarity  100.0%;  Pred. No. 8.6e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy       1 CGCGCGCG 8
           ||||||||
Db       3 CGCGCGCG 10


RESULT 42
AR199370/c
LOCUS       AR199370                  12 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 79 from patent US 6355428.
ACCESSION   AR199370
VERSION     AR199370.1  GI:20249444
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 12)
  AUTHORS   Schroth,G.P., Bruice,T.Wayne. and Suh,Y.J.
  TITLE     Nucleic acid ligand interaction assays
  JOURNAL   Patent: US 6355428-A 79 12-MAR-2002;
FEATURES             Location/Qualifiers
     source          1. .12
                     /organism="unknown"
BASE COUNT        0 a      8 c      4 g      0 t
ORIGIN

  Query Match             100.0%;  Score 8;  DB 6;  Length 12;
  Best Local Similarity   100.0%;  Pred. No. 8.6e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db      10 CGCGCGCG 3


RESULT 43
AR048207
LOCUS       AR048207                  16 bp    DNA     linear   PAT 29-SEP-1999
DEFINITION  Sequence 3 from patent US 5821070.
ACCESSION   AR048207
VERSION     AR048207.1  GI:5970550
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 16)
  AUTHORS   Lee,W.-H. and Shan,B.
  TITLE     Antibodies reactive with retinoblastoma binding proteins and
            methods of using same
  JOURNAL   Patent: US 5821070-A 3 13-OCT-1998;
FEATURES             Location/Qualifiers
     source          1. .16
                     /organism="unknown"
BASE COUNT        3 a      5 c      5 g      3 t
ORIGIN

  Query Match             100.0%;  Score 8;  DB 6;  Length 16;
  Best Local Similarity   100.0%;  Pred. No. 8.1e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       6 CGCGCGCG 13


RESULT 44
AR048207/c
LOCUS       AR048207                  16 bp    DNA     linear   PAT 29-SEP-1999
DEFINITION  Sequence 3 from patent US 5821070.
ACCESSION   AR048207
VERSION     AR048207.1  GI:5970550
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 16)
  AUTHORS   Lee,W.-H. and Shan,B.
  TITLE     Antibodies reactive with retinoblastoma binding proteins and
            methods of using same
  JOURNAL   Patent: US 5821070-A 3 13-OCT-1998;
FEATURES             Location/Qualifiers
     source          1. .16
                     /organism="unknown"
BASE COUNT        3 a      5 c      5 g      3 t
ORIGIN

  Query Match             100.0%;  Score 8;  DB 6;  Length 16;
  Best Local Similarity   100.0%;  Pred. No. 8.1e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db      11 CGCGCGCG 4


RESULT 45
AR055680
LOCUS       AR055680                  16 bp    DNA     linear   PAT 29-SEP-1999
DEFINITION  Sequence 3 from patent US 5837512.
ACCESSION   AR055680
VERSION     AR055680.1  GI:5981257
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 16)
  AUTHORS   Rabson,A.B., Lin,H.-C., Bodkin,M. and Strair,R.
  TITLE     HIV-1 vectors
  JOURNAL   Patent: US 5837512-A 3 17-NOV-1998;
FEATURES             Location/Qualifiers
     source          1. .16
                     /organism="unknown"
BASE COUNT        0 a      8 c      8 g      0 t
ORIGIN

  Query Match             100.0%;  Score 8;  DB 6;  Length 16;
  Best Local Similarity   100.0%;  Pred. No. 8.1e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       8 CGCGCGCG 15


Search completed: March 14, 2003, 04:08:03
Job time : 1364 secs
```

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: March 14, 2003, 01:45:32 ; Search time 197 Seconds
(without alignments)
91.452 Million cell updates/sec

Title: CGCGCGCG
Perfect score: 8
Sequence: 1 cgcgcgcg 8

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	8	100.0	8	22	AAS11907	Hypothetical forwa
c	2	8	100.0	8	22	AAS11907	Hypothetical forwa
	3	8	100.0	10	20	AAX77475	US5912147 primer 1
c	4	8	100.0	10	20	AAX77475	US5912147 primer 1
	5	8	100.0	10	20	AAX77476	US5912147 primer 2
c	6	8	100.0	10	20	AAX77476	US5912147 primer 2
	7	8	100.0	10	20	AAX77477	US5912147 primer 2
c	8	8	100.0	10	20	AAX77477	US5912147 primer 2
	9	8	100.0	10	20	AAX77478	US5912147 primer 2
c	10	8	100.0	10	20	AAX77478	US5912147 primer 2
	11	8	100.0	10	20	AAX77479	US5912147 primer 2
c	12	8	100.0	10	20	AAX77479	US5912147 primer 2
	13	8	100.0	10	20	AAX77481	US5912147 primer 2
c	14	8	100.0	10	20	AAX77481	US5912147 primer 2
	15	8	100.0	10	20	AAX77480	US5912147 primer 2
c	16	8	100.0	10	20	AAX77480	US5912147 primer 2
	17	8	100.0	10	20	AAX77482	US5912147 primer 2
c	18	8	100.0	10	20	AAX77482	US5912147 primer 2
	19	8	100.0	10	20	AAX77483	US5912147 primer 2
c	20	8	100.0	10	20	AAX77483	US5912147 primer 2
	21	8	100.0	10	20	AAX77465	US5912147 primer 9
c	22	8	100.0	10	20	AAX77465	US5912147 primer 9
	23	8	100.0	10	20	AAX77466	US5912147 primer 1
c	24	8	100.0	10	20	AAX77466	US5912147 primer 1
	25	8	100.0	10	20	AAX77467	US5912147 primer 1
c	26	8	100.0	10	20	AAX77467	US5912147 primer 1
	27	8	100.0	10	20	AAX77468	US5912147 primer 1
c	28	8	100.0	10	20	AAX77468	US5912147 primer 1
	29	8	100.0	10	20	AAX77469	US5912147 primer 1
c	30	8	100.0	10	20	AAX77469	US5912147 primer 1
	31	8	100.0	10	20	AAX77470	US5912147 primer 1
c	32	8	100.0	10	20	AAX77470	US5912147 primer 1
	33	8	100.0	10	20	AAX77471	US5912147 primer 1
c	34	8	100.0	10	20	AAX77471	US5912147 primer 1
	35	8	100.0	10	20	AAX77472	US5912147 primer 1
c	36	8	100.0	10	20	AAX77472	US5912147 primer 1
	37	8	100.0	10	20	AAX77473	US5912147 primer 1
c	38	8	100.0	10	20	AAX77473	US5912147 primer 1
	39	8	100.0	10	20	AAX77474	US5912147 primer 1
c	40	8	100.0	10	20	AAX77474	US5912147 primer 1
	41	8	100.0	12	21	AAA10396	Kinetic strand dis
c	42	8	100.0	12	21	AAA10396	Kinetic strand dis
	43	8	100.0	12	23	AAS13767	Simple sequence re
c	44	8	100.0	12	23	AAS13767	Simple sequence re
	45	8	100.0	12	23	ABH86501	Oligonucleotide pr

ALIGNMENTS

RESULT 1
AAS11907
ID AAS11907 standard; DNA; 8 BP.
XX
AC AAS11907;
XX
DT 07-NOV-2001 (first entry)
XX
DE Hypothetical forward PCR primer.
XX
KW Hypothetical PCR primer; DNA sequencing; cDNA cloning;
KW single-stranded DNA amplification; circular template; ss.
XX
OS Synthetic.
XX
PN WO200159101-A1.
XX
PD 16-AUG-2001.
XX
PF 09-FEB-2001; 2001WO-US04259.
XX
PR 10-FEB-2000; 2000US-0181615.
PR 09-MAY-2000; 2000US-0203035.
XX
PA (PENN-) PENN STATE RES FOUND.
XX
PI Connor JR, Ye Z;
XX
DR WPI; 2001-522475/57.
XX
PT Amplifying a polynucleotide sequence for molecular biology protocols,
PT comprises circularizing a linear, single strand polynucleotide sample

```
PT   and initiating primer extension amplification using sequence specific
PT   primers  -
XX
PS   Disclosure; Page 8; 39pp; English.
XX
CC   The invention relates to methods of amplifying a polynucleotide sequence,
CC   involving obtaining a linear, single strand polynucleotide sample,
CC   ligating the ends of the sample to form a circular shaped sample,
CC   introducing first and second sequence specific primers to the circular
CC   sample and initiating a primer extension amplification reaction to
CC   increase copy number of the circular sample. The new methods are useful
CC   for amplifying a polynucleotide sequence, cloning a full length cDNA
CC   sequence from an mRNA sample and sequencing a full length coding DNA or
CC   mRNA for a gene (cloned). The products obtained are useful for a number
CC   of molecular biology protocols including diagnostics, sequencing and
CC   mutation. The method uses a single-strand of polynucleotides and can
CC   amplify the single strand cDNA obtained after reverse transcription of
CC   mRNA rather than double-stranded cDNA, hence the method is accurate and
CC   efficient in amplification. The method is less time consuming and
CC   inexpensive. Background of polymerase chain reaction (PCR) products is
CC   reduced as specific primers are used in all amplification reactions and
CC   as the primer binds to both ends of the cDNA, and even low expressed
CC   mRNA can be cloned. Since the method uses only first strand cDNA as a
CC   PCR template, the longest first strand cDNA could be synthesised by
CC   using reverse transcriptase without RNase H activity. The cDNA band of
CC   correct size can be obtained on the first pass or the amplification of
CC   cDNA ends can be repeated until the correct size of the cDNA is
CC   obtained. The amplification of cDNA towards the ends, which is contrary
CC   to normal gene structure, decreases the possibility of contamination in
CC   cDNA cloning from genomic DNA. The technique can also be used as a
CC   better alternative to existing methods for 5' end primer extension
CC   because of its ability to specifically amplify cDNA ends using a grade
CC   series of amplification steps. The present sequence is a hypothetical
CC   PCR primer used to illustrate the method of the invention.
XX
SQ   Sequence 8 BP; 0 A; 4 C; 4 G; 0 U; 0 other;

  Query Match             100.0%;  Score 8;  DB 22;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 2.8e+08;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       1 CGCGCGCG 8


RESULT 2
AAS11907/c
ID   AAS11907 standard; DNA; 8 BP.
XX
AC   AAS11907;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Hypothetical forward PCR primer.
XX
KW   Hypothetical PCR primer; DNA sequencing; cDNA cloning;
KW   single-stranded DNA amplification; circular template; ss.
XX
OS   Synthetic.
XX
PN   WO200159101-A1.
XX
PD   16-AUG-2001.
XX
PF   09-FEB-2001; 2001WO-US04259.
XX
PR   10-FEB-2000; 2000US-0181615.
PR   09-MAY-2000; 2000US-0203035.
XX
PA   (PENN-) PENN STATE RES FOUND.
XX
PI   Connor JR,  Ye Z;
XX
DR   WPI; 2001-522475/57.
XX
PT   Amplifying a polynucleotide sequence for molecular biology protocols,
PT   comprises circularizing a linear, single strand polynucleotide sample
PT   and initiating primer extension amplification using sequence specific
PT   primers  -
XX
PS   Disclosure; Page 8; 39pp; English.
XX
CC   The invention relates to methods of amplifying a polynucleotide sequence,
CC   involving obtaining a linear, single strand polynucleotide sample,
CC   ligating the ends of the sample to form a circular shaped sample,
CC   introducing first and second sequence specific primers to the circular
CC   sample and initiating a primer extension amplification reaction to
CC   increase copy number of the circular sample. The new methods are useful
CC   for amplifying a polynucleotide sequence, cloning a full length cDNA
CC   sequence from an mRNA sample and sequencing a full length coding DNA or
CC   mRNA for a gene (cloned). The products obtained are useful for a number
CC   of molecular biology protocols including diagnostics, sequencing and
CC   mutation. The method uses a single-strand of polynucleotides and can
CC   amplify the single strand cDNA obtained after reverse transcription of
CC   mRNA rather than double-stranded cDNA, hence the method is accurate and
CC   efficient in amplification. The method is less time consuming and
CC   inexpensive. Background of polymerase chain reaction (PCR) products is
CC   reduced as specific primers are used in all amplification reactions and
CC   as the primer binds to both ends of the cDNA, and even low expressed
CC   mRNA can be cloned. Since the method uses only first strand cDNA as a
CC   PCR template, the longest first strand cDNA could be synthesised by
CC   using reverse transcriptase without RNase H activity. The cDNA band of
CC   correct size can be obtained on the first pass or the amplification of
CC   cDNA ends can be repeated until the correct size of the cDNA is
CC   obtained. The amplification of cDNA towards the ends, which is contrary
CC   to normal gene structure, decreases the possibility of contamination in
CC   cDNA cloning from genomic DNA. The technique can also be used as a
CC   better alternative to existing methods for 5' end primer extension
CC   because of its ability to specifically amplify cDNA ends using a grade
CC   series of amplification steps. The present sequence is a hypothetical
CC   PCR primer used to illustrate the method of the invention.
XX
SQ   Sequence 8 BP; 0 A; 4 C; 4 G; 0 U; 0 other;

  Query Match             100.0%;  Score 8;  DB 22;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 2.8e+08;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       8 CGCGCGCG 1


RESULT 3
AAX77475
ID   AAX77475 standard; DNA; 10 BP.
XX
AC   AAX77475;
XX
DT   05-AUG-1999  (first entry)
XX
DE   US5912147 primer 19.
XX
KW   Primer; quantitation; genetic instability; tumour cell; detection;
KW   neoplastic transformation; carcinogenesis; ss.
XX
OS   Synthetic.
XX
PN   US5912147-A.
XX
PD   15-JUN-1999.
XX
PF   22-OCT-1996;   96US-0734973.
XX
PR   22-OCT-1996;   96US-0734973.
```

XX
PA (HEAL-) HEALTH RES INC.
XX
PI Anderson G, Basik M, Stoler D;
XX
DR WPI; 1999-357197/30.
XX
PT Quantitating genetic instability
XX
PS Claim 4; Column 23-24; 27pp; English.
XX
CC This invention describes a novel method for quantitating genetic
CC instability independent of microsatellite alterations by treating a
CC comparison pair comprising genomic DNA from tumour cells and genomic DNA
CC from normal cells. The method involves the cells from the same individual
CC with oligonucleotide primers selected from (i) a nucleotide sequence
CC (CG)xRG, where R is a purine selected from adenine and guanine and x =
CC 3-7, (ii) a nucleotide sequence (CG)xRY, where R is as in (i) and Y is a
CC pyrimidine selected from cytosine, thymine, and uracil and x = 3-7, (iii)
CC a nucleotide sequence (CG)xRR, where R is as in (i) and x = 3-7, (iv) a
CC nucleotide sequence (CG)xYY, where Y is a pyrimidine selected from
CC cytosine, thymine, and uracil and x = 3-7, (v) a nucleotide sequence
CC (CA)xRG, where R is a purine selected from adenine and guanine and x =
CC 6-16, (vi) a nucleotide sequence (CA)xRY, where R is a purine selected
CC from adenine and guanine and Y is a pyrimidine selected from cytosine,
CC thymine, and uracil, and x = 6-16, (vii) a nucleotide sequence (CA)xRR,
CC where R is a purine selected from adenine and guanine and x = 6-16,
CC (viii) a nucleotide sequence (CA)xYY, where Y is a pyrimidine selected
CC from cytosine, thymine, and uracil and x = 6-16, and (ix) a combination
CC of the primers. The method is useful for detecting genomic instability
CC which are commonly associated with the various stages of neoplastic
CC transformation and carcinogenesis. The method is rapid and simple.
XX
SQ Sequence 10 BP; 0 A; 6 C; 4 G; 0 U; 0 other;

Query Match 100.0%; Score 8; DB 20; Length 10;
Best Local Similarity 100.0%; Pred. No. 5.3e+04;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGCGCGCG 8
||||||||
Db 1 CGCGCGCG 8

RESULT 4
AAX77475/c
ID AAX77475 standard; DNA; 10 BP.
XX
AC AAX77475;
XX
DT 05-AUG-1999 (first entry)
XX
DE US5912147 primer 19.
XX
KW Primer; quantitation; genetic instability; tumour cell; detection;
KW neoplastic transformation; carcinogenesis; ss.
XX
OS Synthetic.
XX
PN US5912147-A.
XX
PD 15-JUN-1999.
XX
PF 22-OCT-1996; 96US-0734973.
XX
PR 22-OCT-1996; 96US-0734973.
XX
PA (HEAL-) HEALTH RES INC.
XX
PI Anderson G, Basik M, Stoler D;
XX
DR WPI; 1999-357197/30.
XX
PT Quantitating genetic instability
XX
PS Claim 4; Column 23-24; 27pp; English.
XX
CC This invention describes a novel method for quantitating genetic
CC instability independent of microsatellite alterations by treating a
CC comparison pair comprising genomic DNA from tumour cells and genomic DNA
CC from normal cells. The method involves the cells from the same individual
CC with oligonucleotide primers selected from (i) a nucleotide sequence
CC (CG)xRG, where R is a purine selected from adenine and guanine and x =
CC 3-7, (ii) a nucleotide sequence (CG)xRY, where R is as in (i) and Y is a
CC pyrimidine selected from cytosine, thymine, and uracil and x = 3-7, (iii)
CC a nucleotide sequence (CG)xRR, where R is as in (i) and x = 3-7, (iv) a
CC nucleotide sequence (CG)xYY, where Y is a pyrimidine selected from
CC cytosine, thymine, and uracil and x = 3-7, (v) a nucleotide sequence
CC (CA)xRG, where R is a purine selected from adenine and guanine and x =
CC 6-16, (vi) a nucleotide sequence (CA)xRY, where R is a purine selected
CC from adenine and guanine and Y is a pyrimidine selected from cytosine,
CC thymine, and uracil, and x = 6-16, (vii) a nucleotide sequence (CA)xRR,
CC where R is a purine selected from adenine and guanine and x = 6-16,
CC (viii) a nucleotide sequence (CA)xYY, where Y is a pyrimidine selected
CC from cytosine, thymine, and uracil and x = 6-16, and (ix) a combination
CC of the primers. The method is useful for detecting genomic instability
CC which are commonly associated with the various stages of neoplastic
CC transformation and carcinogenesis. The method is rapid and simple.
XX
SQ Sequence 10 BP; 0 A; 6 C; 4 G; 0 U; 0 other;

Query Match 100.0%; Score 8; DB 20; Length 10;
Best Local Similarity 100.0%; Pred. No. 5.3e+04;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGCGCGCG 8
||||||||
Db 8 CGCGCGCG 1

RESULT 5
AAX77476
ID AAX77476 standard; DNA; 10 BP.
XX
AC AAX77476;
XX
DT 05-AUG-1999 (first entry)
XX
DE US5912147 primer 20.
XX
KW Primer; quantitation; genetic instability; tumour cell; detection;
KW neoplastic transformation; carcinogenesis; ss.
XX
OS Synthetic.
XX
PN US5912147-A.
XX
PD 15-JUN-1999.
XX
PF 22-OCT-1996; 96US-0734973.
XX
PR 22-OCT-1996; 96US-0734973.
XX
PA (HEAL-) HEALTH RES INC.
XX
PI Anderson G, Basik M, Stoler D;
XX
DR WPI; 1999-357197/30.
XX
PT Quantitating genetic instability
XX
PS Claim 4; Column 23-24; 27pp; English.
XX
CC This invention describes a novel method for quantitating genetic
CC instability independent of microsatellite alterations by treating a
CC comparison pair comprising genomic DNA from tumour cells and genomic DNA

```
CC   from normal cells. The method involves the cells from the same individual
CC   with oligonucleotide primers selected from (i) a nucleotide sequence
CC   (CG)xRG, where R is a purine selected from adenine and guanine and x =
CC   3-7, (ii) a nucleotide sequence (CG)xRY, where R is as in (i) and Y is a
CC   pyrimidine selected from cytosine, thymine, and uracil and x = 3-7, (iii)
CC   a nucleotide sequence (CG)xRR, where R is as in (i) and x = 3-7, (iv) a
CC   nucleotide sequence (CG)xYY, where Y is a pyrimidine selected from
CC   cytosine, thymine, and uracil and x = 3-7, (v) a nucleotide sequence
CC   (CA)xRG, where R is a purine selected from adenine and guanine and x =
CC   6-16, (vi) a nucleotide sequence (CA)xRY, where R is a purine selected
CC   from adenine and guanine and Y is a pyrimidine  selected from cytosine,
CC   thymine, and uracil,  and x = 6-16, (vii) a nucleotide sequence (CA)xRR,
CC   where R is a purine selected from adenine and guanine and x = 6-16,
CC   (viii) a nucleotide sequence (CA)xYY, where Y is a pyrimidine selected
CC   from cytosine, thymine, and uracil and x = 6-16, and (ix) a combination
CC   of the primers. The method is useful for detecting genomic instability
CC   which are commonly associated with the various stages of neoplastic
CC   transformation and carcinogenesis. The method is rapid and simple.
XX
SQ   Sequence 10 BP; 0 A; 5 C; 4 G; 1 T; 0 other;

  Query Match             100.0%;  Score 8;  DB 20;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 5.3e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       1 CGCGCGCG 8


RESULT 6
AAX77476/c
ID   AAX77476 standard; DNA; 10 BP.
XX
AC   AAX77476;
XX
DT   05-AUG-1999  (first entry)
XX
DE   US5912147 primer 20.
XX
KW   Primer; quantitation; genetic instability; tumour cell; detection;
KW   neoplastic transformation; carcinogenesis; ss.
XX
OS   Synthetic.
XX
PN   US5912147-A.
XX
PD   15-JUN-1999.
XX
PF   22-OCT-1996;   96US-0734973.
XX
PR   22-OCT-1996;   96US-0734973.
XX
PA   (HEAL-) HEALTH RES INC.
XX
PI   Anderson G,  Basik M,  Stoler D;
XX
DR   WPI; 1999-357197/30.
XX
PT   Quantitating genetic instability
XX
PS   Claim 4; Column 23-24; 27pp; English.
XX
CC   This invention describes a novel method for quantitating genetic
CC   instability independent of microsatellite alterations by treating a
CC   comparison pair comprising genomic DNA from tumour cells and genomic DNA
CC   from normal cells. The method involves the cells from the same individual
CC   with oligonucleotide primers selected from (i) a nucleotide sequence
CC   (CG)xRG, where R is a purine selected from adenine and guanine and x =
CC   3-7, (ii) a nucleotide sequence (CG)xRY, where R is as in (i) and Y is a
CC   pyrimidine selected from cytosine, thymine, and uracil and x = 3-7, (iii)
CC   a nucleotide sequence (CG)xRR, where R is as in (i) and x = 3-7, (iv) a
CC   nucleotide sequence (CG)xYY, where Y is a pyrimidine selected from
CC   cytosine, thymine, and uracil and x = 3-7, (v) a nucleotide sequence
CC   (CA)xRG, where R is a purine selected from adenine and guanine and x =
CC   6-16, (vi) a nucleotide sequence (CA)xRY, where R is a purine selected
CC   from adenine and guanine and Y is a pyrimidine  selected from cytosine,
CC   thymine, and uracil,  and x = 6-16, (vii) a nucleotide sequence (CA)xRR,
CC   where R is a purine selected from adenine and guanine and x = 6-16,
CC   (viii) a nucleotide sequence (CA)xYY, where Y is a pyrimidine selected
CC   from cytosine, thymine, and uracil and x = 6-16, and (ix) a combination
CC   of the primers. The method is useful for detecting genomic instability
CC   which are commonly associated with the various stages of neoplastic
CC   transformation and carcinogenesis. The method is rapid and simple.
XX
SQ   Sequence 10 BP; 0 A; 5 C; 4 G; 1 T; 0 other;

  Query Match             100.0%;  Score 8;  DB 20;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 5.3e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       8 CGCGCGCG 1


RESULT 7
AAX77477
ID   AAX77477 standard; DNA; 10 BP.
XX
AC   AAX77477;
XX
DT   05-AUG-1999  (first entry)
XX
DE   US5912147 primer 21.
XX
KW   Primer; quantitation; genetic instability; tumour cell; detection;
KW   neoplastic transformation; carcinogenesis; DNA/RNA hybrid; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   misc_RNA        10
FT                   /*tag=  a
FT                   /note= "uracil"
XX
PN   US5912147-A.
XX
PD   15-JUN-1999.
XX
PF   22-OCT-1996;   96US-0734973.
XX
PR   22-OCT-1996;   96US-0734973.
XX
PA   (HEAL-) HEALTH RES INC.
XX
PI   Anderson G,  Basik M,  Stoler D;
XX
DR   WPI; 1999-357197/30.
XX
PT   Quantitating genetic instability
XX
PS   Claim 4; Column 25-26; 27pp; English.
XX
CC   This invention describes a novel method for quantitating genetic
CC   instability independent of microsatellite alterations by treating a
CC   comparison pair comprising genomic DNA from tumour cells and genomic DNA
CC   from normal cells. The method involves the cells from the same individual
CC   with oligonucleotide primers selected from (i) a nucleotide sequence
CC   (CG)xRG, where R is a purine selected from adenine and guanine and x =
CC   3-7, (ii) a nucleotide sequence (CG)xRY, where R is as in (i) and Y is a
CC   pyrimidine selected from cytosine, thymine, and uracil and x = 3-7, (iii)
CC   a nucleotide sequence (CG)xRR, where R is as in (i) and x = 3-7, (iv) a
CC   nucleotide sequence (CG)xYY, where Y is a pyrimidine selected from
CC   cytosine, thymine, and uracil and x = 3-7, (v) a nucleotide sequence
CC   (CA)xRG, where R is a purine selected from adenine and guanine and x =
```

```
CC   6-16, (vi) a nucleotide sequence (CA)xRY, where R is a purine selected
CC   from adenine and guanine and Y is a pyrimidine  selected from cytosine,
CC   thymine, and uracil,  and x = 6-16, (vii) a nucleotide sequence (CA)xRR,
CC   where R is a purine selected from adenine and guanine and x = 6-16,
CC   (viii) a nucleotide sequence (CA)xYY, where Y is a pyrimidine selected
CC   from cytosine, thymine, and uracil and x = 6-16, and (ix) a combination
CC   of the primers. The method is useful for detecting genomic instability
CC   which are commonly associated with the various stages of neoplastic
CC   transformation and carcinogenesis. The method is rapid and simple.
XX
SQ   Sequence 10 BP; 0 A; 5 C; 4 G; 1 U; 0 other;

  Query Match           100.0%;  Score 8;  DB 20;  Length 10;
  Best Local Similarity  100.0%;  Pred. No. 5.3e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db      1 CGCGCGCG 8


RESULT 8
AAX77477/c
ID   AAX77477 standard; DNA; 10 BP.
XX
AC   AAX77477;
XX
DT   05-AUG-1999  (first entry)
XX
DE   US5912147 primer 21.
XX
KW   Primer; quantitation; genetic instability; tumour cell; detection;
KW   neoplastic transformation; carcinogenesis; DNA/RNA hybrid; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   misc_RNA        10
FT                   /*tag= a
FT                   /note= "uracil"
XX
PN   US5912147-A.
XX
PD   15-JUN-1999.
XX
PF   22-OCT-1996;   96US-0734973.
XX
PR   22-OCT-1996;   96US-0734973.
XX
PA   (HEAL-) HEALTH RES INC.
XX
PI   Anderson G,  Basik M,  Stoler D;
XX
DR   WPI; 1999-357197/30.
XX
PT   Quantitating genetic instability
XX
PS   Claim 4; Column 25-26; 27pp; English.
XX
CC   This invention describes a novel method for quantitating genetic
CC   instability independent of microsatellite alterations by treating a
CC   comparison pair comprising genomic DNA from tumour cells and genomic DNA
CC   from normal cells. The method involves the cells from the same individual
CC   with oligonucleotide primers selected from (i) a nucleotide sequence
CC   (CG)xRG, where R is a purine selected from adenine and guanine and x =
CC   3-7, (ii) a nucleotide sequence (CG)xRY, where R is as in (i) and Y is a
CC   pyrimidine selected from cytosine, thymine, and uracil and x = 3-7, (iii)
CC   a nucleotide sequence (CG)xRR, where R is as in (i) and x = 3-7, (iv) a
CC   nucleotide sequence (CG)xYY, where Y is a pyrimidine selected from
CC   cytosine, thymine, and uracil and x = 3-7, (v) a nucleotide sequence
CC   (CA)xRG, where R is a purine selected from adenine and guanine and x =
CC   6-16, (vi) a nucleotide sequence (CA)xRY, where R is a purine selected
CC   from adenine and guanine and Y is a pyrimidine  selected from cytosine,
CC   thymine, and uracil,  and x = 6-16, (vii) a nucleotide sequence (CA)xRR,
CC   where R is a purine selected from adenine and guanine and x = 6-16,
CC   (viii) a nucleotide sequence (CA)xYY, where Y is a pyrimidine selected
CC   from cytosine, thymine, and uracil and x = 6-16, and (ix) a combination
CC   of the primers. The method is useful for detecting genomic instability
CC   which are commonly associated with the various stages of neoplastic
CC   transformation and carcinogenesis. The method is rapid and simple.
XX
SQ   Sequence 10 BP; 0 A; 5 C; 4 G; 1 U; 0 other;

  Query Match           100.0%;  Score 8;  DB 20;  Length 10;
  Best Local Similarity  100.0%;  Pred. No. 5.3e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db      8 CGCGCGCG 1


RESULT 9
AAX77478
ID   AAX77478 standard; DNA; 10 BP.
XX
AC   AAX77478;
XX
DT   05-AUG-1999  (first entry)
XX
DE   US5912147 primer 22.
XX
KW   Primer; quantitation; genetic instability; tumour cell; detection;
KW   neoplastic transformation; carcinogenesis; ss.
XX
OS   Synthetic.
XX
PN   US5912147-A.
XX
PD   15-JUN-1999.
XX
PF   22-OCT-1996;   96US-0734973.
XX
PR   22-OCT-1996;   96US-0734973.
XX
PA   (HEAL-) HEALTH RES INC.
XX
PI   Anderson G,  Basik M,  Stoler D;
XX
DR   WPI; 1999-357197/30.
XX
PT   Quantitating genetic instability
XX
PS   Claim 4; Column 25-26; 27pp; English.
XX
CC   This invention describes a novel method for quantitating genetic
CC   instability independent of microsatellite alterations by treating a
CC   comparison pair comprising genomic DNA from tumour cells and genomic DNA
CC   from normal cells. The method involves the cells from the same individual
CC   with oligonucleotide primers selected from (i) a nucleotide sequence
CC   (CG)xRG, where R is a purine selected from adenine and guanine and x =
CC   3-7, (ii) a nucleotide sequence (CG)xRY, where R is as in (i) and Y is a
CC   pyrimidine selected from cytosine, thymine, and uracil and x = 3-7, (iii)
CC   a nucleotide sequence (CG)xRR, where R is as in (i) and x = 3-7, (iv) a
CC   nucleotide sequence (CG)xYY, where Y is a pyrimidine selected from
CC   cytosine, thymine, and uracil and x = 3-7, (v) a nucleotide sequence
CC   (CA)xRG, where R is a purine selected from adenine and guanine and x =
CC   6-16, (vi) a nucleotide sequence (CA)xRY, where R is a purine selected
CC   from adenine and guanine and Y is a pyrimidine  selected from cytosine,
CC   thymine, and uracil,  and x = 6-16, (vii) a nucleotide sequence (CA)xRR,
CC   where R is a purine selected from adenine and guanine and x = 6-16,
CC   (viii) a nucleotide sequence (CA)xYY, where Y is a pyrimidine selected
CC   from cytosine, thymine, and uracil and x = 6-16, and (ix) a combination
CC   of the primers. The method is useful for detecting genomic instability
CC   which are commonly associated with the various stages of neoplastic
CC   transformation and carcinogenesis. The method is rapid and simple.
```

```
XX
SQ   Sequence 10 BP; 0 A; 4 C; 4 G; 2 T; 0 other;

  Query Match            100.0%;  Score 8;  DB 20;  Length 10;
  Best Local Similarity  100.0%;  Pred. No. 5.3e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       1 CGCGCGCG 8


RESULT 10
AAX77478/c
ID   AAX77478 standard; DNA; 10 BP.
XX
AC   AAX77478;
XX
DT   05-AUG-1999  (first entry)
XX
DE   US5912147 primer 22.
XX
KW   Primer; quantitation; genetic instability; tumour cell; detection;
KW   neoplastic transformation; carcinogenesis; ss.
XX
OS   Synthetic.
XX
PN   US5912147-A.
XX
PD   15-JUN-1999.
XX
PF   22-OCT-1996;   96US-0734973.
XX
PR   22-OCT-1996;   96US-0734973.
XX
PA   (HEAL-) HEALTH RES INC.
XX
PI   Anderson G,  Basik M,  Stoler D;
XX
DR   WPI; 1999-357197/30.
XX
PT   Quantitating genetic instability
XX
PS   Claim 4; Column 25-26; 27pp; English.
XX
CC   This invention describes a novel method for quantitating genetic
CC   instability independent of microsatellite alterations by treating a
CC   comparison pair comprising genomic DNA from tumour cells and genomic DNA
CC   from normal cells. The method involves the cells from the same individual
CC   with oligonucleotide primers selected from (i) a nucleotide sequence
CC   (CG)xRG, where R is a purine selected from adenine and guanine and x =
CC   3-7, (ii) a nucleotide sequence (CG)xRY, where R is as in (i) and Y is a
CC   pyrimidine selected from cytosine, thymine, and uracil and x = 3-7, (iii)
CC   a nucleotide sequence (CG)xRR, where R is as in (i) and x = 3-7, (iv) a
CC   nucleotide sequence (CG)xYY, where Y is a pyrimidine selected from
CC   cytosine, thymine, and uracil and x = 3-7, (v) a nucleotide sequence
CC   (CA)xRG, where R is a purine selected from adenine and guanine and x =
CC   6-16, (vi) a nucleotide sequence (CA)xRY, where R is a purine selected
CC   from adenine and guanine and Y is a pyrimidine  selected from cytosine,
CC   thymine, and uracil,  and x = 6-16, (vii) a nucleotide sequence (CA)xRR,
CC   where R is a purine selected from adenine and guanine and x = 6-16,
CC   (viii) a nucleotide sequence (CA)xYY, where Y is a pyrimidine selected
CC   from cytosine, thymine, and uracil and x = 6-16, and (ix) a combination
CC   of the primers. The method is useful for detecting genomic instability
CC   which are commonly associated with the various stages of neoplastic
CC   transformation and carcinogenesis. The method is rapid and simple.
XX
SQ   Sequence 10 BP; 0 A; 4 C; 4 G; 2 T; 0 other;

  Query Match            100.0%;  Score 8;  DB 20;  Length 10;
  Best Local Similarity  100.0%;  Pred. No. 5.3e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       8 CGCGCGCG 1


RESULT 11
AAX77479
ID   AAX77479 standard; DNA; 10 BP.
XX
AC   AAX77479;
XX
DT   05-AUG-1999  (first entry)
XX
DE   US5912147 primer 23.
XX
KW   Primer; quantitation; genetic instability; tumour cell; detection;
KW   neoplastic transformation; carcinogenesis; ss.
XX
OS   Synthetic.
XX
PN   US5912147-A.
XX
PD   15-JUN-1999.
XX
PF   22-OCT-1996;   96US-0734973.
XX
PR   22-OCT-1996;   96US-0734973.
XX
PA   (HEAL-) HEALTH RES INC.
XX
PI   Anderson G,  Basik M,  Stoler D;
XX
DR   WPI; 1999-357197/30.
XX
PT   Quantitating genetic instability
XX
PS   Claim 4; Column 25-26; 27pp; English.
XX
CC   This invention describes a novel method for quantitating genetic
CC   instability independent of microsatellite alterations by treating a
CC   comparison pair comprising genomic DNA from tumour cells and genomic DNA
CC   from normal cells. The method involves the cells from the same individual
CC   with oligonucleotide primers selected from (i) a nucleotide sequence
CC   (CG)xRG, where R is a purine selected from adenine and guanine and x =
CC   3-7, (ii) a nucleotide sequence (CG)xRY, where R is as in (i) and Y is a
CC   pyrimidine selected from cytosine, thymine, and uracil and x = 3-7, (iii)
CC   a nucleotide sequence (CG)xRR, where R is as in (i) and x = 3-7, (iv) a
CC   nucleotide sequence (CG)xYY, where Y is a pyrimidine selected from
CC   cytosine, thymine, and uracil and x = 3-7, (v) a nucleotide sequence
CC   (CA)xRG, where R is a purine selected from adenine and guanine and x =
CC   6-16, (vi) a nucleotide sequence (CA)xRY, where R is a purine selected
CC   from adenine and guanine and Y is a pyrimidine  selected from cytosine,
CC   thymine, and uracil,  and x = 6-16, (vii) a nucleotide sequence (CA)xRR,
CC   where R is a purine selected from adenine and guanine and x = 6-16,
CC   (viii) a nucleotide sequence (CA)xYY, where Y is a pyrimidine selected
CC   from cytosine, thymine, and uracil and x = 6-16, and (ix) a combination
CC   of the primers. The method is useful for detecting genomic instability
CC   which are commonly associated with the various stages of neoplastic
CC   transformation and carcinogenesis. The method is rapid and simple.
XX
SQ   Sequence 10 BP; 0 A; 5 C; 4 G; 1 T; 0 other;

  Query Match            100.0%;  Score 8;  DB 20;  Length 10;
  Best Local Similarity  100.0%;  Pred. No. 5.3e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       1 CGCGCGCG 8


RESULT 12
AAX77479/c
```

```
ID   AAX77479 standard; DNA; 10 BP.
XX
AC   AAX77479;
XX
DT   05-AUG-1999  (first entry)
XX
DE   US5912147 primer 23.
XX
KW   Primer; quantitation; genetic instability; tumour cell; detection;
KW   neoplastic transformation; carcinogenesis; ss.
XX
OS   Synthetic.
XX
PN   US5912147-A.
XX
PD   15-JUN-1999.
XX
PF   22-OCT-1996;   96US-0734973.
XX
PR   22-OCT-1996;   96US-0734973.
XX
PA   (HEAL-) HEALTH RES INC.
XX
PI   Anderson G,  Basik M,  Stoler D;
XX
DR   WPI; 1999-357197/30.
XX
PT   Quantitating genetic instability
XX
PS   Claim 4; Column 25-26; 27pp; English.
XX
CC   This invention describes a novel method for quantitating genetic
CC   instability independent of microsatellite alterations by treating a
CC   comparison pair comprising genomic DNA from tumour cells and genomic DNA
CC   from normal cells. The method involves the cells from the same individual
CC   with oligonucleotide primers selected from (i) a nucleotide sequence
CC   (CG)xRG, where R is a purine selected from adenine and guanine and x =
CC   3-7, (ii) a nucleotide sequence (CG)xRY, where R is as in (i) and Y is a
CC   pyrimidine selected from cytosine, thymine, and uracil and x = 3-7, (iii)
CC   a nucleotide sequence (CG)xRR, where R is as in (i) and x = 3-7, (iv) a
CC   nucleotide sequence (CG)xYY, where Y is a pyrimidine selected from
CC   cytosine, thymine, and uracil and x = 3-7, (v) a nucleotide sequence
CC   (CA)xRG, where R is a purine selected from adenine and guanine and x =
CC   6-16, (vi) a nucleotide sequence (CA)xRY, where R is a purine selected
CC   from adenine and guanine and Y is a pyrimidine  selected from cytosine,
CC   thymine, and uracil,  and x = 6-16, (vii) a nucleotide sequence (CA)xRR,
CC   where R is a purine selected from adenine and guanine and x = 6-16,
CC   (viii) a nucleotide sequence (CA)xYY, where Y is a pyrimidine selected
CC   from cytosine, thymine, and uracil and x = 6-16, and (ix) a combination
CC   of the primers. The method is useful for detecting genomic instability
CC   which are commonly associated with the various stages of neoplastic
CC   transformation and carcinogenesis. The method is rapid and simple.
XX
SQ   Sequence 10 BP; 0 A; 5 C; 4 G; 1 T; 0 other;

  Query Match             100.0%;  Score 8;  DB 20;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 5.3e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       8 CGCGCGCG 1


RESULT 13
AAX77481
ID   AAX77481 standard; DNA; 10 BP.
XX
AC   AAX77481;
XX
DT   05-AUG-1999  (first entry)
XX
DE   US5912147 primer 25.
XX
KW   Primer; quantitation; genetic instability; tumour cell; detection;
KW   neoplastic transformation; carcinogenesis; DNA/RNA hybrid; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   misc_RNA        9
FT                   /*tag= a
FT                   /note= "uracil"
XX
PN   US5912147-A.
XX
PD   15-JUN-1999.
XX
PF   22-OCT-1996;   96US-0734973.
XX
PR   22-OCT-1996;   96US-0734973.
XX
PA   (HEAL-) HEALTH RES INC.
XX
PI   Anderson G,  Basik M,  Stoler D;
XX
DR   WPI; 1999-357197/30.
XX
PT   Quantitating genetic instability
XX
PS   Claim 4; Column 25-26; 27pp; English.
XX
CC   This invention describes a novel method for quantitating genetic
CC   instability independent of microsatellite alterations by treating a
CC   comparison pair comprising genomic DNA from tumour cells and genomic DNA
CC   from normal cells. The method involves the cells from the same individual
CC   with oligonucleotide primers selected from (i) a nucleotide sequence
CC   (CG)xRG, where R is a purine selected from adenine and guanine and x =
CC   3-7, (ii) a nucleotide sequence (CG)xRY, where R is as in (i) and Y is a
CC   pyrimidine selected from cytosine, thymine, and uracil and x = 3-7, (iii)
CC   a nucleotide sequence (CG)xRR, where R is as in (i) and x = 3-7, (iv) a
CC   nucleotide sequence (CG)xYY, where Y is a pyrimidine selected from
CC   cytosine, thymine, and uracil and x = 3-7, (v) a nucleotide sequence
CC   (CA)xRG, where R is a purine selected from adenine and guanine and x =
CC   6-16, (vi) a nucleotide sequence (CA)xRY, where R is a purine selected
CC   from adenine and guanine and Y is a pyrimidine  selected from cytosine,
CC   thymine, and uracil,  and x = 6-16, (vii) a nucleotide sequence (CA)xRR,
CC   where R is a purine selected from adenine and guanine and x = 6-16,
CC   (viii) a nucleotide sequence (CA)xYY, where Y is a pyrimidine selected
CC   from cytosine, thymine, and uracil and x = 6-16, and (ix) a combination
CC   of the primers. The method is useful for detecting genomic instability
CC   which are commonly associated with the various stages of neoplastic
CC   transformation and carcinogenesis. The method is rapid and simple.
XX
SQ   Sequence 10 BP; 0 A; 5 C; 4 G; 1 U; 0 other;

  Query Match             100.0%;  Score 8;  DB 20;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 5.3e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       1 CGCGCGCG 8


RESULT 14
AAX77481/c
ID   AAX77481 standard; DNA; 10 BP.
XX
AC   AAX77481;
XX
DT   05-AUG-1999  (first entry)
XX
DE   US5912147 primer 25.
XX
KW   Primer; quantitation; genetic instability; tumour cell; detection;
```

```
KW   neoplastic transformation; carcinogenesis; DNA/RNA hybrid; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   misc_RNA        9
FT                   /*tag=  a
FT                   /note= "uracil"
XX
PN   US5912147-A.
XX
PD   15-JUN-1999.
XX
PF   22-OCT-1996;    96US-0734973.
XX
PR   22-OCT-1996;    96US-0734973.
XX
PA   (HEAL-) HEALTH RES INC.
XX
PI   Anderson G,  Basik M,  Stoler D;
XX
DR   WPI; 1999-357197/30.
XX
PT   Quantitating genetic instability
XX
PS   Claim 4; Column 25-26; 27pp; English.
XX
CC   This invention describes a novel method for quantitating genetic
CC   instability independent of microsatellite alterations by treating a
CC   comparison pair comprising genomic DNA from tumour cells and genomic DNA
CC   from normal cells. The method involves the cells from the same individual
CC   with oligonucleotide primers selected from (i) a nucleotide sequence
CC   (CG)xRG, where R is a purine selected from adenine and guanine and x =
CC   3-7, (ii) a nucleotide sequence (CG)xRY, where R is as in (i) and Y is a
CC   pyrimidine selected from cytosine, thymine, and uracil and x = 3-7, (iii)
CC   a nucleotide sequence (CG)xRR, where R is as in (i) and x = 3-7, (iv) a
CC   nucleotide sequence (CG)xYY, where Y is a pyrimidine selected from
CC   cytosine, thymine, and uracil and x = 3-7, (v) a nucleotide sequence
CC   (CA)xRG, where R is a purine selected from adenine and guanine and x =
CC   6-16, (vi) a nucleotide sequence (CA)xRY, where R is a purine selected
CC   from adenine and guanine and Y is a pyrimidine  selected from cytosine,
CC   thymine, and uracil,  and x = 6-16, (vii) a nucleotide sequence (CA)xRR,
CC   where R is a purine selected from adenine and guanine and x = 6-16,
CC   (viii) a nucleotide sequence (CA)xYY, where Y is a pyrimidine selected
CC   from cytosine, thymine, and uracil and x = 6-16, and (ix) a combination
CC   of the primers. The method is useful for detecting genomic instability
CC   which are commonly associated with the various stages of neoplastic
CC   transformation and carcinogenesis. The method is rapid and simple.
XX
SQ   Sequence 10 BP; 0 A; 5 C; 4 G; 1 U; 0 other;

  Query Match             100.0%;  Score 8;  DB 20;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 5.3e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       8 CGCGCGCG 1


RESULT 15
AAX77480
ID   AAX77480 standard; DNA; 10 BP.
XX
AC   AAX77480;
XX
DT   05-AUG-1999  (first entry)
XX
DE   US5912147 primer 24.
XX
KW   Primer; quantitation; genetic instability; tumour cell; detection;
KW   neoplastic transformation; carcinogenesis; DNA/RNA hybrid; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   misc_RNA        10
FT                   /*tag=  a
FT                   /note= "uracil"
XX
PN   US5912147-A.
XX
PD   15-JUN-1999.
XX
PF   22-OCT-1996;    96US-0734973.
XX
PR   22-OCT-1996;    96US-0734973.
XX
PA   (HEAL-) HEALTH RES INC.
XX
PI   Anderson G,  Basik M,  Stoler D;
XX
DR   WPI; 1999-357197/30.
XX
PT   Quantitating genetic instability
XX
PS   Claim 4; Column 25-26; 27pp; English.
XX
CC   This invention describes a novel method for quantitating genetic
CC   instability independent of microsatellite alterations by treating a
CC   comparison pair comprising genomic DNA from tumour cells and genomic DNA
CC   from normal cells. The method involves the cells from the same individual
CC   with oligonucleotide primers selected from (i) a nucleotide sequence
CC   (CG)xRG, where R is a purine selected from adenine and guanine and x =
CC   3-7, (ii) a nucleotide sequence (CG)xRY, where R is as in (i) and Y is a
CC   pyrimidine selected from cytosine, thymine, and uracil and x = 3-7, (iii)
CC   a nucleotide sequence (CG)xRR, where R is as in (i) and x = 3-7, (iv) a
CC   nucleotide sequence (CG)xYY, where Y is a pyrimidine selected from
CC   cytosine, thymine, and uracil and x = 3-7, (v) a nucleotide sequence
CC   (CA)xRG, where R is a purine selected from adenine and guanine and x =
CC   6-16, (vi) a nucleotide sequence (CA)xRY, where R is a purine selected
CC   from adenine and guanine and Y is a pyrimidine  selected from cytosine,
CC   thymine, and uracil,  and x = 6-16, (vii) a nucleotide sequence (CA)xRR,
CC   where R is a purine selected from adenine and guanine and x = 6-16,
CC   (viii) a nucleotide sequence (CA)xYY, where Y is a pyrimidine selected
CC   from cytosine, thymine, and uracil and x = 6-16, and (ix) a combination
CC   of the primers. The method is useful for detecting genomic instability
CC   which are commonly associated with the various stages of neoplastic
CC   transformation and carcinogenesis. The method is rapid and simple.
XX
SQ   Sequence 10 BP; 0 A; 4 C; 4 G; 1 T; 1 U; 0 other;

  Query Match             100.0%;  Score 8;  DB 20;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 5.3e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       1 CGCGCGCG 8


RESULT 16
AAX77480/c
ID   AAX77480 standard; DNA; 10 BP.
XX
AC   AAX77480;
XX
DT   05-AUG-1999  (first entry)
XX
DE   US5912147 primer 24.
XX
KW   Primer; quantitation; genetic instability; tumour cell; detection;
KW   neoplastic transformation; carcinogenesis; DNA/RNA hybrid; ss.
XX
OS   Synthetic.
XX
```

```
FH   Key             Location/Qualifiers
FT   misc_RNA        10
FT                   /*tag= a
FT                   /note= "uracil"
XX
PN   US5912147-A.
XX
PD   15-JUN-1999.
XX
PF   22-OCT-1996;    96US-0734973.
XX
PR   22-OCT-1996;    96US-0734973.
XX
PA   (HEAL-) HEALTH RES INC.
XX
PI   Anderson G,  Basik M,  Stoler D;
XX
DR   WPI; 1999-357197/30.
XX
PT   Quantitating genetic instability
XX
PS   Claim 4; Column 25-26; 27pp; English.
XX
CC   This invention describes a novel method for quantitating genetic
CC   instability independent of microsatellite alterations by treating a
CC   comparison pair comprising genomic DNA from tumour cells and genomic DNA
CC   from normal cells. The method involves the cells from the same individual
CC   with oligonucleotide primers selected from (i) a nucleotide sequence
CC   (CG)xRG, where R is a purine selected from adenine and guanine and x =
CC   3-7, (ii) a nucleotide sequence (CG)xRY, where R is as in (i) and Y is a
CC   pyrimidine selected from cytosine, thymine, and uracil and x = 3-7, (iii)
CC   a nucleotide sequence (CG)xRR, where R is as in (i) and x = 3-7, (iv) a
CC   nucleotide sequence (CG)xYY, where Y is a pyrimidine selected from
CC   cytosine, thymine, and uracil and x = 3-7, (v) a nucleotide sequence
CC   (CA)xRG, where R is a purine selected from adenine and guanine and x =
CC   6-16, (vi) a nucleotide sequence (CA)xRY, where R is a purine selected
CC   from adenine and guanine and Y is a pyrimidine  selected from cytosine,
CC   thymine, and uracil,  and x = 6-16, (vii) a nucleotide sequence (CA)xRR,
CC   where R is a purine selected from adenine and guanine and x = 6-16,
CC   (viii) a nucleotide sequence (CA)xYY, where Y is a pyrimidine selected
CC   from cytosine, thymine, and uracil and x = 6-16, and (ix) a combination
CC   of the primers. The method is useful for detecting genomic instability
CC   which are commonly associated with the various stages of neoplastic
CC   transformation and carcinogenesis. The method is rapid and simple.
XX
SQ   Sequence 10 BP; 0 A; 4 C; 4 G; 1 T; 1 U; 0 other;

  Query Match           100.0%;  Score 8;  DB 20;  Length 10;
  Best Local Similarity  100.0%;  Pred. No. 5.3e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       8 CGCGCGCG 1


RESULT 17
AAX77482
ID   AAX77482 standard; DNA; 10 BP.
XX
AC   AAX77482;
XX
DT   05-AUG-1999  (first entry)
XX
DE   US5912147 primer 26.
XX
KW   Primer; quantitation; genetic instability; tumour cell; detection;
KW   neoplastic transformation; carcinogenesis; DNA/RNA hybrid; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   misc_RNA        9
FT                   /*tag= a
FT                   /note= "uracil"
XX
PN   US5912147-A.
XX
PD   15-JUN-1999.
XX
PF   22-OCT-1996;    96US-0734973.
XX
PR   22-OCT-1996;    96US-0734973.
XX
PA   (HEAL-) HEALTH RES INC.
XX
PI   Anderson G,  Basik M,  Stoler D;
XX
DR   WPI; 1999-357197/30.
XX
PT   Quantitating genetic instability
XX
PS   Claim 4; Column 27-28; 27pp; English.
XX
CC   This invention describes a novel method for quantitating genetic
CC   instability independent of microsatellite alterations by treating a
CC   comparison pair comprising genomic DNA from tumour cells and genomic DNA
CC   from normal cells. The method involves the cells from the same individual
CC   with oligonucleotide primers selected from (i) a nucleotide sequence
CC   (CG)xRG, where R is a purine selected from adenine and guanine and x =
CC   3-7, (ii) a nucleotide sequence (CG)xRY, where R is as in (i) and Y is a
CC   pyrimidine selected from cytosine, thymine, and uracil and x = 3-7, (iii)
CC   a nucleotide sequence (CG)xRR, where R is as in (i) and x = 3-7, (iv) a
CC   nucleotide sequence (CG)xYY, where Y is a pyrimidine selected from
CC   cytosine, thymine, and uracil and x = 3-7, (v) a nucleotide sequence
CC   (CA)xRG, where R is a purine selected from adenine and guanine and x =
CC   6-16, (vi) a nucleotide sequence (CA)xRY, where R is a purine selected
CC   from adenine and guanine and Y is a pyrimidine  selected from cytosine,
CC   thymine, and uracil,  and x = 6-16, (vii) a nucleotide sequence (CA)xRR,
CC   where R is a purine selected from adenine and guanine and x = 6-16,
CC   (viii) a nucleotide sequence (CA)xYY, where Y is a pyrimidine selected
CC   from cytosine, thymine, and uracil and x = 6-16, and (ix) a combination
CC   of the primers. The method is useful for detecting genomic instability
CC   which are commonly associated with the various stages of neoplastic
CC   transformation and carcinogenesis. The method is rapid and simple.
XX
SQ   Sequence 10 BP; 0 A; 4 C; 4 G; 1 T; 1 U; 0 other;

  Query Match           100.0%;  Score 8;  DB 20;  Length 10;
  Best Local Similarity  100.0%;  Pred. No. 5.3e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       1 CGCGCGCG 8


RESULT 18
AAX77482/c
ID   AAX77482 standard; DNA; 10 BP.
XX
AC   AAX77482;
XX
DT   05-AUG-1999  (first entry)
XX
DE   US5912147 primer 26.
XX
KW   Primer; quantitation; genetic instability; tumour cell; detection;
KW   neoplastic transformation; carcinogenesis; DNA/RNA hybrid; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   misc_RNA        9
FT                   /*tag= a
FT                   /note= "uracil"
```

```
XX
PN   US5912147-A.
XX
PD   15-JUN-1999.
XX
PF   22-OCT-1996;   96US-0734973.
XX
PR   22-OCT-1996;   96US-0734973.
XX
PA   (HEAL-) HEALTH RES INC.
XX
PI   Anderson G,  Basik M,  Stoler D;
XX
DR   WPI; 1999-357197/30.
XX
PT   Quantitating genetic instability
XX
PS   Claim 4; Column 27-28; 27pp; English.
XX
CC   This invention describes a novel method for quantitating genetic
CC   instability independent of microsatellite alterations by treating a
CC   comparison pair comprising genomic DNA from tumour cells and genomic DNA
CC   from normal cells. The method involves the cells from the same individual
CC   with oligonucleotide primers selected from (i) a nucleotide sequence
CC   (CG)xRG, where R is a purine selected from adenine and guanine and x =
CC   3-7, (ii) a nucleotide sequence (CG)xRY, where R is as in (i) and Y is a
CC   pyrimidine selected from cytosine, thymine, and uracil and x = 3-7, (iii)
CC   a nucleotide sequence (CG)xRR, where R is as in (i) and x = 3-7, (iv) a
CC   nucleotide sequence (CG)xYY, where Y is a pyrimidine selected from
CC   cytosine, thymine, and uracil and x = 3-7, (v) a nucleotide sequence
CC   (CA)xRG, where R is a purine selected from adenine and guanine and x =
CC   6-16, (vi) a nucleotide sequence (CA)xRY, where R is a purine selected
CC   from adenine and guanine and Y is a pyrimidine  selected from cytosine,
CC   thymine, and uracil,  and x = 6-16, (vii) a nucleotide sequence (CA)xRR,
CC   where R is a purine selected from adenine and guanine and x = 6-16,
CC   (viii) a nucleotide sequence (CA)xYY, where Y is a pyrimidine selected
CC   from cytosine, thymine, and uracil and x = 6-16, and (ix) a combination
CC   of the primers. The method is useful for detecting genomic instability
CC   which are commonly associated with the various stages of neoplastic
CC   transformation and carcinogenesis. The method is rapid and simple.
XX
SQ   Sequence 10 BP; 0 A; 4 C; 4 G; 1 T; 1 U; 0 other;

  Query Match             100.0%;  Score 8;  DB 20;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 5.3e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       8 CGCGCGCG 1


RESULT 19
AAX77483
ID   AAX77483 standard; DNA; 10 BP.
XX
AC   AAX77483;
XX
DT   05-AUG-1999  (first entry)
XX
DE   US5912147 primer 27.
XX
KW   Primer; quantitation; genetic instability; tumour cell; detection;
KW   neoplastic transformation; carcinogenesis; DNA/RNA hybrid; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   misc_RNA        9..10
FT                   /*tag=  a
FT                   /note= "uracil"
XX
PN   US5912147-A.
XX
PD   15-JUN-1999.
XX
PF   22-OCT-1996;   96US-0734973.
XX
PR   22-OCT-1996;   96US-0734973.
XX
PA   (HEAL-) HEALTH RES INC.
XX
PI   Anderson G,  Basik M,  Stoler D;
XX
DR   WPI; 1999-357197/30.
XX
PT   Quantitating genetic instability
XX
PS   Claim 4; Column 27-28; 27pp; English.
XX
CC   This invention describes a novel method for quantitating genetic
CC   instability independent of microsatellite alterations by treating a
CC   comparison pair comprising genomic DNA from tumour cells and genomic DNA
CC   from normal cells. The method involves the cells from the same individual
CC   with oligonucleotide primers selected from (i) a nucleotide sequence
CC   (CG)xRG, where R is a purine selected from adenine and guanine and x =
CC   3-7, (ii) a nucleotide sequence (CG)xRY, where R is as in (i) and Y is a
CC   pyrimidine selected from cytosine, thymine, and uracil and x = 3-7, (iii)
CC   a nucleotide sequence (CG)xRR, where R is as in (i) and x = 3-7, (iv) a
CC   nucleotide sequence (CG)xYY, where Y is a pyrimidine selected from
CC   cytosine, thymine, and uracil and x = 3-7, (v) a nucleotide sequence
CC   (CA)xRG, where R is a purine selected from adenine and guanine and x =
CC   6-16, (vi) a nucleotide sequence (CA)xRY, where R is a purine selected
CC   from adenine and guanine and Y is a pyrimidine  selected from cytosine,
CC   thymine, and uracil,  and x = 6-16, (vii) a nucleotide sequence (CA)xRR,
CC   where R is a purine selected from adenine and guanine and x = 6-16,
CC   (viii) a nucleotide sequence (CA)xYY, where Y is a pyrimidine selected
CC   from cytosine, thymine, and uracil and x = 6-16, and (ix) a combination
CC   of the primers. The method is useful for detecting genomic instability
CC   which are commonly associated with the various stages of neoplastic
CC   transformation and carcinogenesis. The method is rapid and simple.
XX
SQ   Sequence 10 BP; 0 A; 4 C; 4 G; 2 U; 0 other;

  Query Match             100.0%;  Score 8;  DB 20;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 5.3e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       1 CGCGCGCG 8


RESULT 20
AAX77483/c
ID   AAX77483 standard; DNA; 10 BP.
XX
AC   AAX77483;
XX
DT   05-AUG-1999  (first entry)
XX
DE   US5912147 primer 27.
XX
KW   Primer; quantitation; genetic instability; tumour cell; detection;
KW   neoplastic transformation; carcinogenesis; DNA/RNA hybrid; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   misc_RNA        9..10
FT                   /*tag=  a
FT                   /note= "uracil"
XX
PN   US5912147-A.
XX
PD   15-JUN-1999.
```

```
XX
PF   22-OCT-1996;   96US-0734973.
XX
PR   22-OCT-1996;   96US-0734973.
XX
PA   (HEAL-) HEALTH RES INC.
XX
PI   Anderson G,  Basik M,  Stoler D;
XX
DR   WPI; 1999-357197/30.
XX
PT   Quantitating genetic instability
XX
PS   Claim 4; Column 27-28; 27pp; English.
XX
CC   This invention describes a novel method for quantitating genetic
CC   instability independent of microsatellite alterations by treating a
CC   comparison pair comprising genomic DNA from tumour cells and genomic DNA
CC   from normal cells. The method involves the cells from the same individual
CC   with oligonucleotide primers selected from (i) a nucleotide sequence
CC   (CG)xRG, where R is a purine selected from adenine and guanine and x =
CC   3-7, (ii) a nucleotide sequence (CG)xRY, where R is as in (i) and Y is a
CC   pyrimidine selected from cytosine, thymine, and uracil and x = 3-7, (iii)
CC   a nucleotide sequence (CG)xRR, where R is as in (i) and x = 3-7, (iv) a
CC   nucleotide sequence (CG)xYY, where Y is a pyrimidine selected from
CC   cytosine, thymine, and uracil and x = 3-7, (v) a nucleotide sequence
CC   (CA)xRG, where R is a purine selected from adenine and guanine and x =
CC   6-16, (vi) a nucleotide sequence (CA)xRY, where R is a purine selected
CC   from adenine and guanine and Y is a pyrimidine  selected from cytosine,
CC   thymine, and uracil,  and x = 6-16, (vii) a nucleotide sequence (CA)xRR,
CC   where R is a purine selected from adenine and guanine and x = 6-16,
CC   (viii) a nucleotide sequence (CA)xYY, where Y is a pyrimidine selected
CC   from cytosine, thymine, and uracil and x = 6-16, and (ix) a combination
CC   of the primers. The method is useful for detecting genomic instability
CC   which are commonly associated with the various stages of neoplastic
CC   transformation and carcinogenesis. The method is rapid and simple.
XX
SQ   Sequence 10 BP; 0 A; 4 C; 4 G; 2 U; 0 other;

  Query Match             100.0%;  Score 8;  DB 20;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 5.3e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       8 CGCGCGCG 1


RESULT 21
AAX77465
ID   AAX77465 standard; DNA; 10 BP.
XX
AC   AAX77465;
XX
DT   05-AUG-1999  (first entry)
XX
DE   US5912147 primer 9.
XX
KW   Primer; quantitation; genetic instability; tumour cell; detection;
KW   neoplastic transformation; carcinogenesis; ss.
XX
OS   Synthetic.
XX
PN   US5912147-A.
XX
PD   15-JUN-1999.
XX
PF   22-OCT-1996;   96US-0734973.
XX
PR   22-OCT-1996;   96US-0734973.
XX
PA   (HEAL-) HEALTH RES INC.
XX
PI   Anderson G,  Basik M,  Stoler D;
XX
DR   WPI; 1999-357197/30.
XX
PT   Quantitating genetic instability
XX
PS   Claim 4; Column 19-20; 27pp; English.
XX
CC   This invention describes a novel method for quantitating genetic
CC   instability independent of microsatellite alterations by treating a
CC   comparison pair comprising genomic DNA from tumour cells and genomic DNA
CC   from normal cells. The method involves the cells from the same individual
CC   with oligonucleotide primers selected from (i) a nucleotide sequence
CC   (CG)xRG, where R is a purine selected from adenine and guanine and x =
CC   3-7, (ii) a nucleotide sequence (CG)xRY, where R is as in (i) and Y is a
CC   pyrimidine selected from cytosine, thymine, and uracil and x = 3-7, (iii)
CC   a nucleotide sequence (CG)xRR, where R is as in (i) and x = 3-7, (iv) a
CC   nucleotide sequence (CG)xYY, where Y is a pyrimidine selected from
CC   cytosine, thymine, and uracil and x = 3-7, (v) a nucleotide sequence
CC   (CA)xRG, where R is a purine selected from adenine and guanine and x =
CC   6-16, (vi) a nucleotide sequence (CA)xRY, where R is a purine selected
CC   from adenine and guanine and Y is a pyrimidine  selected from cytosine,
CC   thymine, and uracil,  and x = 6-16, (vii) a nucleotide sequence (CA)xRR,
CC   where R is a purine selected from adenine and guanine and x = 6-16,
CC   (viii) a nucleotide sequence (CA)xYY, where Y is a pyrimidine selected
CC   from cytosine, thymine, and uracil and x = 6-16, and (ix) a combination
CC   of the primers. The method is useful for detecting genomic instability
CC   which are commonly associated with the various stages of neoplastic
CC   transformation and carcinogenesis. The method is rapid and simple.
XX
SQ   Sequence 10 BP; 1 A; 4 C; 5 G; 0 U; 0 other;

  Query Match             100.0%;  Score 8;  DB 20;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 5.3e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       1 CGCGCGCG 8


RESULT 22
AAX77465/c
ID   AAX77465 standard; DNA; 10 BP.
XX
AC   AAX77465;
XX
DT   05-AUG-1999  (first entry)
XX
DE   US5912147 primer 9.
XX
KW   Primer; quantitation; genetic instability; tumour cell; detection;
KW   neoplastic transformation; carcinogenesis; ss.
XX
OS   Synthetic.
XX
PN   US5912147-A.
XX
PD   15-JUN-1999.
XX
PF   22-OCT-1996;   96US-0734973.
XX
PR   22-OCT-1996;   96US-0734973.
XX
PA   (HEAL-) HEALTH RES INC.
XX
PI   Anderson G,  Basik M,  Stoler D;
XX
DR   WPI; 1999-357197/30.
XX
PT   Quantitating genetic instability
XX
PS   Claim 4; Column 19-20; 27pp; English.
```

```
XX
CC   This invention describes a novel method for quantitating genetic
CC   instability independent of microsatellite alterations by treating a
CC   comparison pair comprising genomic DNA from tumour cells and genomic DNA
CC   from normal cells. The method involves the cells from the same individual
CC   with oligonucleotide primers selected from (i) a nucleotide sequence
CC   (CG)xRG, where R is a purine selected from adenine and guanine and x =
CC   3-7, (ii) a nucleotide sequence (CG)xRY, where R is as in (i) and Y is a
CC   pyrimidine selected from cytosine, thymine, and uracil and x = 3-7, (iii)
CC   a nucleotide sequence (CG)xRR, where R is as in (i) and x = 3-7, (iv) a
CC   nucleotide sequence (CG)xYY, where Y is a pyrimidine selected from
CC   cytosine, thymine, and uracil and x = 3-7, (v) a nucleotide sequence
CC   (CA)xRG, where R is a purine selected from adenine and guanine and x =
CC   6-16, (vi) a nucleotide sequence (CA)xRY, where R is a purine selected
CC   from adenine and guanine and Y is a pyrimidine  selected from cytosine,
CC   thymine, and uracil,  and x = 6-16, (vii) a nucleotide sequence (CA)xRR,
CC   where R is a purine selected from adenine and guanine and x = 6-16,
CC   (viii) a nucleotide sequence (CA)xYY, where Y is a pyrimidine selected
CC   from cytosine, thymine, and uracil and x = 6-16, and (ix) a combination
CC   of the primers. The method is useful for detecting genomic instability
CC   which are commonly associated with the various stages of neoplastic
CC   transformation and carcinogenesis. The method is rapid and simple.
XX
SQ   Sequence 10 BP; 1 A; 4 C; 5 G; 0 U; 0 other;

  Query Match             100.0%;  Score 8;  DB 20;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 5.3e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       8 CGCGCGCG 1


RESULT 23
AAX77466
ID   AAX77466 standard; DNA; 10 BP.
XX
AC   AAX77466;
XX
DT   05-AUG-1999  (first entry)
XX
DE   US5912147 primer 10.
XX
KW   Primer; quantitation; genetic instability; tumour cell; detection;
KW   neoplastic transformation; carcinogenesis; ss.
XX
OS   Synthetic.
XX
PN   US5912147-A.
XX
PD   15-JUN-1999.
XX
PF   22-OCT-1996;   96US-0734973.
XX
PR   22-OCT-1996;   96US-0734973.
XX
PA   (HEAL-) HEALTH RES INC.
XX
PI   Anderson G,  Basik M,  Stoler D;
XX
DR   WPI; 1999-357197/30.
XX
PT   Quantitating genetic instability
XX
PS   Claim 4; Column 19-20; 27pp; English.
XX
CC   This invention describes a novel method for quantitating genetic
CC   instability independent of microsatellite alterations by treating a
CC   comparison pair comprising genomic DNA from tumour cells and genomic DNA
CC   from normal cells. The method involves the cells from the same individual
CC   with oligonucleotide primers selected from (i) a nucleotide sequence
CC   (CG)xRG, where R is a purine selected from adenine and guanine and x =
CC   3-7, (ii) a nucleotide sequence (CG)xRY, where R is as in (i) and Y is a
CC   pyrimidine selected from cytosine, thymine, and uracil and x = 3-7, (iii)
CC   a nucleotide sequence (CG)xRR, where R is as in (i) and x = 3-7, (iv) a
CC   nucleotide sequence (CG)xYY, where Y is a pyrimidine selected from
CC   cytosine, thymine, and uracil and x = 3-7, (v) a nucleotide sequence
CC   (CA)xRG, where R is a purine selected from adenine and guanine and x =
CC   6-16, (vi) a nucleotide sequence (CA)xRY, where R is a purine selected
CC   from adenine and guanine and Y is a pyrimidine  selected from cytosine,
CC   thymine, and uracil,  and x = 6-16, (vii) a nucleotide sequence (CA)xRR,
CC   where R is a purine selected from adenine and guanine and x = 6-16,
CC   (viii) a nucleotide sequence (CA)xYY, where Y is a pyrimidine selected
CC   from cytosine, thymine, and uracil and x = 6-16, and (ix) a combination
CC   of the primers. The method is useful for detecting genomic instability
CC   which are commonly associated with the various stages of neoplastic
CC   transformation and carcinogenesis. The method is rapid and simple.
XX
SQ   Sequence 10 BP; 0 A; 4 C; 6 G; 0 U; 0 other;

  Query Match             100.0%;  Score 8;  DB 20;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 5.3e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       1 CGCGCGCG 8


RESULT 24
AAX77466/c
ID   AAX77466 standard; DNA; 10 BP.
XX
AC   AAX77466;
XX
DT   05-AUG-1999  (first entry)
XX
DE   US5912147 primer 10.
XX
KW   Primer; quantitation; genetic instability; tumour cell; detection;
KW   neoplastic transformation; carcinogenesis; ss.
XX
OS   Synthetic.
XX
PN   US5912147-A.
XX
PD   15-JUN-1999.
XX
PF   22-OCT-1996;   96US-0734973.
XX
PR   22-OCT-1996;   96US-0734973.
XX
PA   (HEAL-) HEALTH RES INC.
XX
PI   Anderson G,  Basik M,  Stoler D;
XX
DR   WPI; 1999-357197/30.
XX
PT   Quantitating genetic instability
XX
PS   Claim 4; Column 19-20; 27pp; English.
XX
CC   This invention describes a novel method for quantitating genetic
CC   instability independent of microsatellite alterations by treating a
CC   comparison pair comprising genomic DNA from tumour cells and genomic DNA
CC   from normal cells. The method involves the cells from the same individual
CC   with oligonucleotide primers selected from (i) a nucleotide sequence
CC   (CG)xRG, where R is a purine selected from adenine and guanine and x =
CC   3-7, (ii) a nucleotide sequence (CG)xRY, where R is as in (i) and Y is a
CC   pyrimidine selected from cytosine, thymine, and uracil and x = 3-7, (iii)
CC   a nucleotide sequence (CG)xRR, where R is as in (i) and x = 3-7, (iv) a
CC   nucleotide sequence (CG)xYY, where Y is a pyrimidine selected from
CC   cytosine, thymine, and uracil and x = 3-7, (v) a nucleotide sequence
CC   (CA)xRG, where R is a purine selected from adenine and guanine and x =
CC   6-16, (vi) a nucleotide sequence (CA)xRY, where R is a purine selected
```

```
CC   from adenine and guanine and Y is a pyrimidine  selected from cytosine,
CC   thymine, and uracil,  and x = 6-16, (vii) a nucleotide sequence (CA)xRR,
CC   where R is a purine selected from adenine and guanine and x = 6-16,
CC   (viii) a nucleotide sequence (CA)xYY, where Y is a pyrimidine selected
CC   from cytosine, thymine, and uracil and x = 6-16, and (ix) a combination
CC   of the primers. The method is useful for detecting genomic instability
CC   which are commonly associated with the various stages of neoplastic
CC   transformation and carcinogenesis. The method is rapid and simple.
XX
SQ   Sequence 10 BP; 0 A; 4 C; 6 G; 0 U; 0 other;

  Query Match             100.0%;  Score 8;  DB 20;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 5.3e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db      8 CGCGCGCG 1


RESULT 25
AAX77467
ID   AAX77467 standard; DNA; 10 BP.
XX
AC   AAX77467;
XX
DT   05-AUG-1999  (first entry)
XX
DE   US5912147 primer 11.
XX
KW   Primer; quantitation; genetic instability; tumour cell; detection;
KW   neoplastic transformation; carcinogenesis; ss.
XX
OS   Synthetic.
XX
PN   US5912147-A.
XX
PD   15-JUN-1999.
XX
PF   22-OCT-1996;   96US-0734973.
XX
PR   22-OCT-1996;   96US-0734973.
XX
PA   (HEAL-) HEALTH RES INC.
XX
PI   Anderson G,  Basik M,  Stoler D;
XX
DR   WPI; 1999-357197/30.
XX
PT   Quantitating genetic instability
XX
PS   Claim 4; Column 19-20; 27pp; English.
XX
CC   This invention describes a novel method for quantitating genetic
CC   instability independent of microsatellite alterations by treating a
CC   comparison pair comprising genomic DNA from tumour cells and genomic DNA
CC   from normal cells. The method involves the cells from the same individual
CC   with oligonucleotide primers selected from (i) a nucleotide sequence
CC   (CG)xRG, where R is a purine selected from adenine and guanine and x =
CC   3-7, (ii) a nucleotide sequence (CG)xRY, where R is as in (i) and Y is a
CC   pyrimidine selected from cytosine, thymine, and uracil and x = 3-7, (iii)
CC   a nucleotide sequence (CG)xRR, where R is as in (i) and x = 3-7, (iv) a
CC   nucleotide sequence (CG)xYY, where Y is a pyrimidine selected from
CC   cytosine, thymine, and uracil and x = 3-7, (v) a nucleotide sequence
CC   (CA)xRG, where R is a purine selected from adenine and guanine and x =
CC   6-16, (vi) a nucleotide sequence (CA)xRY, where R is a purine selected
CC   from adenine and guanine and Y is a pyrimidine  selected from cytosine,
CC   thymine, and uracil,  and x = 6-16, (vii) a nucleotide sequence (CA)xRR,
CC   where R is a purine selected from adenine and guanine and x = 6-16,
CC   (viii) a nucleotide sequence (CA)xYY, where Y is a pyrimidine selected
CC   from cytosine, thymine, and uracil and x = 6-16, and (ix) a combination
CC   of the primers. The method is useful for detecting genomic instability
CC   which are commonly associated with the various stages of neoplastic
CC   transformation and carcinogenesis. The method is rapid and simple.
XX
SQ   Sequence 10 BP; 1 A; 5 C; 4 G; 0 U; 0 other;

  Query Match             100.0%;  Score 8;  DB 20;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 5.3e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db      1 CGCGCGCG 8


RESULT 26
AAX77467/c
ID   AAX77467 standard; DNA; 10 BP.
XX
AC   AAX77467;
XX
DT   05-AUG-1999  (first entry)
XX
DE   US5912147 primer 11.
XX
KW   Primer; quantitation; genetic instability; tumour cell; detection;
KW   neoplastic transformation; carcinogenesis; ss.
XX
OS   Synthetic.
XX
PN   US5912147-A.
XX
PD   15-JUN-1999.
XX
PF   22-OCT-1996;   96US-0734973.
XX
PR   22-OCT-1996;   96US-0734973.
XX
PA   (HEAL-) HEALTH RES INC.
XX
PI   Anderson G,  Basik M,  Stoler D;
XX
DR   WPI; 1999-357197/30.
XX
PT   Quantitating genetic instability
XX
PS   Claim 4; Column 19-20; 27pp; English.
XX
CC   This invention describes a novel method for quantitating genetic
CC   instability independent of microsatellite alterations by treating a
CC   comparison pair comprising genomic DNA from tumour cells and genomic DNA
CC   from normal cells. The method involves the cells from the same individual
CC   with oligonucleotide primers selected from (i) a nucleotide sequence
CC   (CG)xRG, where R is a purine selected from adenine and guanine and x =
CC   3-7, (ii) a nucleotide sequence (CG)xRY, where R is as in (i) and Y is a
CC   pyrimidine selected from cytosine, thymine, and uracil and x = 3-7, (iii)
CC   a nucleotide sequence (CG)xRR, where R is as in (i) and x = 3-7, (iv) a
CC   nucleotide sequence (CG)xYY, where Y is a pyrimidine selected from
CC   cytosine, thymine, and uracil and x = 3-7, (v) a nucleotide sequence
CC   (CA)xRG, where R is a purine selected from adenine and guanine and x =
CC   6-16, (vi) a nucleotide sequence (CA)xRY, where R is a purine selected
CC   from adenine and guanine and Y is a pyrimidine  selected from cytosine,
CC   thymine, and uracil,  and x = 6-16, (vii) a nucleotide sequence (CA)xRR,
CC   where R is a purine selected from adenine and guanine and x = 6-16,
CC   (viii) a nucleotide sequence (CA)xYY, where Y is a pyrimidine selected
CC   from cytosine, thymine, and uracil and x = 6-16, and (ix) a combination
CC   of the primers. The method is useful for detecting genomic instability
CC   which are commonly associated with the various stages of neoplastic
CC   transformation and carcinogenesis. The method is rapid and simple.
XX
SQ   Sequence 10 BP; 1 A; 5 C; 4 G; 0 U; 0 other;

  Query Match             100.0%;  Score 8;  DB 20;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 5.3e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy      1 CGCGCGCG 8
          ||||||||
Db      8 CGCGCGCG 1


RESULT 27
AAX77468
ID   AAX77468 standard; DNA; 10 BP.
XX
AC   AAX77468;
XX
DT   05-AUG-1999  (first entry)
XX
DE   US5912147 primer 12.
XX
KW   Primer; quantitation; genetic instability; tumour cell; detection;
KW   neoplastic transformation; carcinogenesis; ss.
XX
OS   Synthetic.
XX
PN   US5912147-A.
XX
PD   15-JUN-1999.
XX
PF   22-OCT-1996;   96US-0734973.
XX
PR   22-OCT-1996;   96US-0734973.
XX
PA   (HEAL-) HEALTH RES INC.
XX
PI   Anderson G,  Basik M,  Stoler D;
XX
DR   WPI; 1999-357197/30.
XX
PT   Quantitating genetic instability
XX
PS   Claim 4; Column 21-22; 27pp; English.
XX
CC   This invention describes a novel method for quantitating genetic
CC   instability independent of microsatellite alterations by treating a
CC   comparison pair comprising genomic DNA from tumour cells and genomic DNA
CC   from normal cells. The method involves the cells from the same individual
CC   with oligonucleotide primers selected from (i) a nucleotide sequence
CC   (CG)xRG, where R is a purine selected from adenine and guanine and x =
CC   3-7, (ii) a nucleotide sequence (CG)xRY, where R is as in (i) and Y is a
CC   pyrimidine selected from cytosine, thymine, and uracil and x = 3-7, (iii)
CC   a nucleotide sequence (CG)xRR, where R is as in (i) and x = 3-7, (iv) a
CC   nucleotide sequence (CG)xYY, where Y is a pyrimidine selected from
CC   cytosine, thymine, and uracil and x = 3-7, (v) a nucleotide sequence
CC   (CA)xRG, where R is a purine selected from adenine and guanine and x =
CC   6-16, (vi) a nucleotide sequence (CA)xRY, where R is a purine selected
CC   from adenine and guanine and Y is a pyrimidine  selected from cytosine,
CC   thymine, and uracil,  and x = 6-16, (vii) a nucleotide sequence (CA)xRR,
CC   where R is a purine selected from adenine and guanine and x = 6-16,
CC   (viii) a nucleotide sequence (CA)xYY, where Y is a pyrimidine selected
CC   from cytosine, thymine, and uracil and x = 6-16, and (ix) a combination
CC   of the primers. The method is useful for detecting genomic instability
CC   which are commonly associated with the various stages of neoplastic
CC   transformation and carcinogenesis. The method is rapid and simple.
XX
SQ   Sequence 10 BP; 1 A; 4 C; 4 G; 1 T; 0 other;

  Query Match             100.0%;  Score 8;  DB 20;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 5.3e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db      1 CGCGCGCG 8


RESULT 28
```

```
AAX77468/c
ID   AAX77468 standard; DNA; 10 BP.
XX
AC   AAX77468;
XX
DT   05-AUG-1999  (first entry)
XX
DE   US5912147 primer 12.
XX
KW   Primer; quantitation; genetic instability; tumour cell; detection;
KW   neoplastic transformation; carcinogenesis; ss.
XX
OS   Synthetic.
XX
PN   US5912147-A.
XX
PD   15-JUN-1999.
XX
PF   22-OCT-1996;   96US-0734973.
XX
PR   22-OCT-1996;   96US-0734973.
XX
PA   (HEAL-) HEALTH RES INC.
XX
PI   Anderson G,  Basik M,  Stoler D;
XX
DR   WPI; 1999-357197/30.
XX
PT   Quantitating genetic instability
XX
PS   Claim 4; Column 21-22; 27pp; English.
XX
CC   This invention describes a novel method for quantitating genetic
CC   instability independent of microsatellite alterations by treating a
CC   comparison pair comprising genomic DNA from tumour cells and genomic DNA
CC   from normal cells. The method involves the cells from the same individual
CC   with oligonucleotide primers selected from (i) a nucleotide sequence
CC   (CG)xRG, where R is a purine selected from adenine and guanine and x =
CC   3-7, (ii) a nucleotide sequence (CG)xRY, where R is as in (i) and Y is a
CC   pyrimidine selected from cytosine, thymine, and uracil and x = 3-7, (iii)
CC   a nucleotide sequence (CG)xRR, where R is as in (i) and x = 3-7, (iv) a
CC   nucleotide sequence (CG)xYY, where Y is a pyrimidine selected from
CC   cytosine, thymine, and uracil and x = 3-7, (v) a nucleotide sequence
CC   (CA)xRG, where R is a purine selected from adenine and guanine and x =
CC   6-16, (vi) a nucleotide sequence (CA)xRY, where R is a purine selected
CC   from adenine and guanine and Y is a pyrimidine  selected from cytosine,
CC   thymine, and uracil,  and x = 6-16, (vii) a nucleotide sequence (CA)xRR,
CC   where R is a purine selected from adenine and guanine and x = 6-16,
CC   (viii) a nucleotide sequence (CA)xYY, where Y is a pyrimidine selected
CC   from cytosine, thymine, and uracil and x = 6-16, and (ix) a combination
CC   of the primers. The method is useful for detecting genomic instability
CC   which are commonly associated with the various stages of neoplastic
CC   transformation and carcinogenesis. The method is rapid and simple.
XX
SQ   Sequence 10 BP; 1 A; 4 C; 4 G; 1 T; 0 other;

  Query Match             100.0%;  Score 8;  DB 20;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 5.3e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db      8 CGCGCGCG 1


RESULT 29
AAX77469
ID   AAX77469 standard; DNA; 10 BP.
XX
AC   AAX77469;
XX
DT   05-AUG-1999  (first entry)
XX
```

```
DE   US5912147 primer 13.
XX
KW   Primer; quantitation; genetic instability; tumour cell; detection;
KW   neoplastic transformation; carcinogenesis; DNA/RNA hybrid; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   misc_RNA        10
FT                   /*tag= a
FT                   /note= "uracil"
XX
PN   US5912147-A.
XX
PD   15-JUN-1999.
XX
PF   22-OCT-1996;    96US-0734973.
XX
PR   22-OCT-1996;    96US-0734973.
XX
PA   (HEAL-) HEALTH RES INC.
XX
PI   Anderson G,  Basik M,  Stoler D;
XX
DR   WPI; 1999-357197/30.
XX
PT   Quantitating genetic instability
XX
PS   Claim 4; Column 21-22; 27pp; English.
XX
CC   This invention describes a novel method for quantitating genetic
CC   instability independent of microsatellite alterations by treating a
CC   comparison pair comprising genomic DNA from tumour cells and genomic DNA
CC   from normal cells. The method involves the cells from the same individual
CC   with oligonucleotide primers selected from (i) a nucleotide sequence
CC   (CG)xRG, where R is a purine selected from adenine and guanine and x =
CC   3-7, (ii) a nucleotide sequence (CG)xRY, where R is as in (i) and Y is a
CC   pyrimidine selected from cytosine, thymine, and uracil and x = 3-7, (iii)
CC   a nucleotide sequence (CG)xRR, where R is as in (i) and x = 3-7, (iv) a
CC   nucleotide sequence (CG)xYY, where Y is a pyrimidine selected from
CC   cytosine, thymine, and uracil and x = 3-7, (v) a nucleotide sequence
CC   (CA)xRG, where R is a purine selected from adenine and guanine and x =
CC   6-16, (vi) a nucleotide sequence (CA)xRY, where R is a purine selected
CC   from adenine and guanine and Y is a pyrimidine  selected from cytosine,
CC   thymine, and uracil,  and x = 6-16, (vii) a nucleotide sequence (CA)xRR,
CC   where R is a purine selected from adenine and guanine and x = 6-16,
CC   (viii) a nucleotide sequence (CA)xYY, where Y is a pyrimidine selected
CC   from cytosine, thymine, and uracil and x = 6-16, and (ix) a combination
CC   of the primers. The method is useful for detecting genomic instability
CC   which are commonly associated with the various stages of neoplastic
CC   transformation and carcinogenesis. The method is rapid and simple.
XX
SQ   Sequence 10 BP; 1 A; 4 C; 4 G; 1 U; 0 other;

  Query Match             100.0%;  Score 8;  DB 20;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 5.3e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       1 CGCGCGCG 8


RESULT 30
AAX77469/c
ID   AAX77469 standard; DNA; 10 BP.
XX
AC   AAX77469;
XX
DT   05-AUG-1999  (first entry)
XX
DE   US5912147 primer 13.
XX
KW   Primer; quantitation; genetic instability; tumour cell; detection;
KW   neoplastic transformation; carcinogenesis; DNA/RNA hybrid; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   misc_RNA        10
FT                   /*tag= a
FT                   /note= "uracil"
XX
PN   US5912147-A.
XX
PD   15-JUN-1999.
XX
PF   22-OCT-1996;    96US-0734973.
XX
PR   22-OCT-1996;    96US-0734973.
XX
PA   (HEAL-) HEALTH RES INC.
XX
PI   Anderson G,  Basik M,  Stoler D;
XX
DR   WPI; 1999-357197/30.
XX
PT   Quantitating genetic instability
XX
PS   Claim 4; Column 21-22; 27pp; English.
XX
CC   This invention describes a novel method for quantitating genetic
CC   instability independent of microsatellite alterations by treating a
CC   comparison pair comprising genomic DNA from tumour cells and genomic DNA
CC   from normal cells. The method involves the cells from the same individual
CC   with oligonucleotide primers selected from (i) a nucleotide sequence
CC   (CG)xRG, where R is a purine selected from adenine and guanine and x =
CC   3-7, (ii) a nucleotide sequence (CG)xRY, where R is as in (i) and Y is a
CC   pyrimidine selected from cytosine, thymine, and uracil and x = 3-7, (iii)
CC   a nucleotide sequence (CG)xRR, where R is as in (i) and x = 3-7, (iv) a
CC   nucleotide sequence (CG)xYY, where Y is a pyrimidine selected from
CC   cytosine, thymine, and uracil and x = 3-7, (v) a nucleotide sequence
CC   (CA)xRG, where R is a purine selected from adenine and guanine and x =
CC   6-16, (vi) a nucleotide sequence (CA)xRY, where R is a purine selected
CC   from adenine and guanine and Y is a pyrimidine  selected from cytosine,
CC   thymine, and uracil,  and x = 6-16, (vii) a nucleotide sequence (CA)xRR,
CC   where R is a purine selected from adenine and guanine and x = 6-16,
CC   (viii) a nucleotide sequence (CA)xYY, where Y is a pyrimidine selected
CC   from cytosine, thymine, and uracil and x = 6-16, and (ix) a combination
CC   of the primers. The method is useful for detecting genomic instability
CC   which are commonly associated with the various stages of neoplastic
CC   transformation and carcinogenesis. The method is rapid and simple.
XX
SQ   Sequence 10 BP; 1 A; 4 C; 4 G; 1 U; 0 other;

  Query Match             100.0%;  Score 8;  DB 20;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 5.3e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       8 CGCGCGCG 1


RESULT 31
AAX77470
ID   AAX77470 standard; DNA; 10 BP.
XX
AC   AAX77470;
XX
DT   05-AUG-1999  (first entry)
XX
DE   US5912147 primer 14.
XX
KW   Primer; quantitation; genetic instability; tumour cell; detection;
KW   neoplastic transformation; carcinogenesis; ss.
```

```
XX
OS   Synthetic.
XX
PN   US5912147-A.
XX
PD   15-JUN-1999.
XX
PF   22-OCT-1996;   96US-0734973.
XX
PR   22-OCT-1996;   96US-0734973.
XX
PA   (HEAL-) HEALTH RES INC.
XX
PI   Anderson G,  Basik M,  Stoler D;
XX
DR   WPI; 1999-357197/30.
XX
PT   Quantitating genetic instability
XX
PS   Claim 4; Column 21-22; 27pp; English.
XX
CC   This invention describes a novel method for quantitating genetic
CC   instability independent of microsatellite alterations by treating a
CC   comparison pair comprising genomic DNA from tumour cells and genomic DNA
CC   from normal cells. The method involves the cells from the same individual
CC   with oligonucleotide primers selected from (i) a nucleotide sequence
CC   (CG)xRG, where R is a purine selected from adenine and guanine and x =
CC   3-7, (ii) a nucleotide sequence (CG)xRY, where R is as in (i) and Y is a
CC   pyrimidine selected from cytosine, thymine, and uracil and x = 3-7, (iii)
CC   a nucleotide sequence (CG)xRR, where R is as in (i) and x = 3-7, (iv) a
CC   nucleotide sequence (CG)xYY, where Y is a pyrimidine selected from
CC   cytosine, thymine, and uracil and x = 3-7, (v) a nucleotide sequence
CC   (CA)xRG, where R is a purine selected from adenine and guanine and x =
CC   6-16, (vi) a nucleotide sequence (CA)xRY, where R is a purine selected
CC   from adenine and guanine and Y is a pyrimidine  selected from cytosine,
CC   thymine, and uracil,  and x = 6-16, (vii) a nucleotide sequence (CA)xRR,
CC   where R is a purine selected from adenine and guanine and x = 6-16,
CC   (viii) a nucleotide sequence (CA)xYY, where Y is a pyrimidine selected
CC   from cytosine, thymine, and uracil and x = 6-16, and (ix) a combination
CC   of the primers. The method is useful for detecting genomic instability
CC   which are commonly associated with the various stages of neoplastic
CC   transformation and carcinogenesis. The method is rapid and simple.
XX
SQ   Sequence 10 BP; 0 A; 5 C; 5 G; 0 U; 0 other;

  Query Match             100.0%;  Score 8;  DB 20;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 5.3e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       1 CGCGCGCG 8


RESULT 32
AAX77470/c
ID   AAX77470 standard; DNA; 10 BP.
XX
AC   AAX77470;
XX
DT   05-AUG-1999  (first entry)
XX
DE   US5912147 primer 14.
XX
KW   Primer; quantitation; genetic instability; tumour cell; detection;
KW   neoplastic transformation; carcinogenesis; ss.
XX
OS   Synthetic.
XX
PN   US5912147-A.
XX
PD   15-JUN-1999.
XX
PF   22-OCT-1996;   96US-0734973.
XX
PR   22-OCT-1996;   96US-0734973.
XX
PA   (HEAL-) HEALTH RES INC.
XX
PI   Anderson G,  Basik M,  Stoler D;
XX
DR   WPI; 1999-357197/30.
XX
PT   Quantitating genetic instability
XX
PS   Claim 4; Column 21-22; 27pp; English.
XX
CC   This invention describes a novel method for quantitating genetic
CC   instability independent of microsatellite alterations by treating a
CC   comparison pair comprising genomic DNA from tumour cells and genomic DNA
CC   from normal cells. The method involves the cells from the same individual
CC   with oligonucleotide primers selected from (i) a nucleotide sequence
CC   (CG)xRG, where R is a purine selected from adenine and guanine and x =
CC   3-7, (ii) a nucleotide sequence (CG)xRY, where R is as in (i) and Y is a
CC   pyrimidine selected from cytosine, thymine, and uracil and x = 3-7, (iii)
CC   a nucleotide sequence (CG)xRR, where R is as in (i) and x = 3-7, (iv) a
CC   nucleotide sequence (CG)xYY, where Y is a pyrimidine selected from
CC   cytosine, thymine, and uracil and x = 3-7, (v) a nucleotide sequence
CC   (CA)xRG, where R is a purine selected from adenine and guanine and x =
CC   6-16, (vi) a nucleotide sequence (CA)xRY, where R is a purine selected
CC   from adenine and guanine and Y is a pyrimidine  selected from cytosine,
CC   thymine, and uracil,  and x = 6-16, (vii) a nucleotide sequence (CA)xRR,
CC   where R is a purine selected from adenine and guanine and x = 6-16,
CC   (viii) a nucleotide sequence (CA)xYY, where Y is a pyrimidine selected
CC   from cytosine, thymine, and uracil and x = 6-16, and (ix) a combination
CC   of the primers. The method is useful for detecting genomic instability
CC   which are commonly associated with the various stages of neoplastic
CC   transformation and carcinogenesis. The method is rapid and simple.
XX
SQ   Sequence 10 BP; 0 A; 5 C; 5 G; 0 U; 0 other;

  Query Match             100.0%;  Score 8;  DB 20;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 5.3e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       8 CGCGCGCG 1


RESULT 33
AAX77471
ID   AAX77471 standard; DNA; 10 BP.
XX
AC   AAX77471;
XX
DT   05-AUG-1999  (first entry)
XX
DE   US5912147 primer 15.
XX
KW   Primer; quantitation; genetic instability; tumour cell; detection;
KW   neoplastic transformation; carcinogenesis; ss.
XX
OS   Synthetic.
XX
PN   US5912147-A.
XX
PD   15-JUN-1999.
XX
PF   22-OCT-1996;   96US-0734973.
XX
PR   22-OCT-1996;   96US-0734973.
XX
PA   (HEAL-) HEALTH RES INC.
XX
PI   Anderson G,  Basik M,  Stoler D;
```

```
XX
DR   WPI; 1999-357197/30.
XX
PT   Quantitating genetic instability
XX
PS   Claim 4; Column 21-22; 27pp; English.
XX
CC   This invention describes a novel method for quantitating genetic
CC   instability independent of microsatellite alterations by treating a
CC   comparison pair comprising genomic DNA from tumour cells and genomic DNA
CC   from normal cells. The method involves the cells from the same individual
CC   with oligonucleotide primers selected from (i) a nucleotide sequence
CC   (CG)xRG, where R is a purine selected from adenine and guanine and x =
CC   3-7, (ii) a nucleotide sequence (CG)xRY, where R is as in (i) and Y is a
CC   pyrimidine selected from cytosine, thymine, and uracil and x = 3-7, (iii)
CC   a nucleotide sequence (CG)xRR, where R is as in (i) and x = 3-7, (iv) a
CC   nucleotide sequence (CG)xYY, where Y is a pyrimidine selected from
CC   cytosine, thymine, and uracil and x = 3-7, (v) a nucleotide sequence
CC   (CA)xRG, where R is a purine selected from adenine and guanine and x =
CC   6-16, (vi) a nucleotide sequence (CA)xRY, where R is a purine selected
CC   from adenine and guanine and Y is a pyrimidine  selected from cytosine,
CC   thymine, and uracil,  and x = 6-16, (vii) a nucleotide sequence (CA)xRR,
CC   where R is a purine selected from adenine and guanine and x = 6-16,
CC   (viii) a nucleotide sequence (CA)xYY, where Y is a pyrimidine selected
CC   from cytosine, thymine, and uracil and x = 6-16, and (ix) a combination
CC   of the primers. The method is useful for detecting genomic instability
CC   which are commonly associated with the various stages of neoplastic
CC   transformation and carcinogenesis. The method is rapid and simple.
XX
SQ   Sequence 10 BP; 0 A; 4 C; 5 G; 1 T; 0 other;

  Query Match             100.0%;  Score 8;  DB 20;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 5.3e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       1 CGCGCGCG 8


RESULT 34
AAX77471/c
ID   AAX77471 standard; DNA; 10 BP.
XX
AC   AAX77471;
XX
DT   05-AUG-1999  (first entry)
XX
DE   US5912147 primer 15.
XX
KW   Primer; quantitation; genetic instability; tumour cell; detection;
KW   neoplastic transformation; carcinogenesis; ss.
XX
OS   Synthetic.
XX
PN   US5912147-A.
XX
PD   15-JUN-1999.
XX
PF   22-OCT-1996;   96US-0734973.
XX
PR   22-OCT-1996;   96US-0734973.
XX
PA   (HEAL-) HEALTH RES INC.
XX
PI   Anderson G,  Basik M,  Stoler D;
XX
DR   WPI; 1999-357197/30.
XX
PT   Quantitating genetic instability
XX
PS   Claim 4; Column 21-22; 27pp; English.
XX
CC   This invention describes a novel method for quantitating genetic
CC   instability independent of microsatellite alterations by treating a
CC   comparison pair comprising genomic DNA from tumour cells and genomic DNA
CC   from normal cells. The method involves the cells from the same individual
CC   with oligonucleotide primers selected from (i) a nucleotide sequence
CC   (CG)xRG, where R is a purine selected from adenine and guanine and x =
CC   3-7, (ii) a nucleotide sequence (CG)xRY, where R is as in (i) and Y is a
CC   pyrimidine selected from cytosine, thymine, and uracil and x = 3-7, (iii)
CC   a nucleotide sequence (CG)xRR, where R is as in (i) and x = 3-7, (iv) a
CC   nucleotide sequence (CG)xYY, where Y is a pyrimidine selected from
CC   cytosine, thymine, and uracil and x = 3-7, (v) a nucleotide sequence
CC   (CA)xRG, where R is a purine selected from adenine and guanine and x =
CC   6-16, (vi) a nucleotide sequence (CA)xRY, where R is a purine selected
CC   from adenine and guanine and Y is a pyrimidine  selected from cytosine,
CC   thymine, and uracil,  and x = 6-16, (vii) a nucleotide sequence (CA)xRR,
CC   where R is a purine selected from adenine and guanine and x = 6-16,
CC   (viii) a nucleotide sequence (CA)xYY, where Y is a pyrimidine selected
CC   from cytosine, thymine, and uracil and x = 6-16, and (ix) a combination
CC   of the primers. The method is useful for detecting genomic instability
CC   which are commonly associated with the various stages of neoplastic
CC   transformation and carcinogenesis. The method is rapid and simple.
XX
SQ   Sequence 10 BP; 0 A; 4 C; 5 G; 1 T; 0 other;

  Query Match             100.0%;  Score 8;  DB 20;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 5.3e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       8 CGCGCGCG 1


RESULT 35
AAX77472
ID   AAX77472 standard; DNA; 10 BP.
XX
AC   AAX77472;
XX
DT   05-AUG-1999  (first entry)
XX
DE   US5912147 primer 16.
XX
KW   Primer; quantitation; genetic instability; tumour cell; detection;
KW   neoplastic transformation; carcinogenesis; DNA/RNA hybrid; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   misc_RNA        10
FT                   /*tag=  a
FT                   /note= "uracil"
XX
PN   US5912147-A.
XX
PD   15-JUN-1999.
XX
PF   22-OCT-1996;   96US-0734973.
XX
PR   22-OCT-1996;   96US-0734973.
XX
PA   (HEAL-) HEALTH RES INC.
XX
PI   Anderson G,  Basik M,  Stoler D;
XX
DR   WPI; 1999-357197/30.
XX
PT   Quantitating genetic instability
XX
PS   Claim 4; Column 23-24; 27pp; English.
XX
CC   This invention describes a novel method for quantitating genetic
CC   instability independent of microsatellite alterations by treating a
```

```
CC   comparison pair comprising genomic DNA from tumour cells and genomic DNA
CC   from normal cells. The method involves the cells from the same individual
CC   with oligonucleotide primers selected from (i) a nucleotide sequence
CC   (CG)xRG, where R is a purine selected from adenine and guanine and x =
CC   3-7, (ii) a nucleotide sequence (CG)xRY, where R is as in (i) and Y is a
CC   pyrimidine selected from cytosine, thymine, and uracil and x = 3-7, (iii)
CC   a nucleotide sequence (CG)xRR, where R is as in (i) and x = 3-7, (iv) a
CC   nucleotide sequence (CG)xYY, where Y is a pyrimidine selected from
CC   cytosine, thymine, and uracil and x = 3-7, (v) a nucleotide sequence
CC   (CA)xRG, where R is a purine selected from adenine and guanine and x =
CC   6-16, (vi) a nucleotide sequence (CA)xRY, where R is a purine selected
CC   from adenine and guanine and Y is a pyrimidine  selected from cytosine,
CC   thymine, and uracil,  and x = 6-16, (vii) a nucleotide sequence (CA)xRR,
CC   where R is a purine selected from adenine and guanine and x = 6-16,
CC   (viii) a nucleotide sequence (CA)xYY, where Y is a pyrimidine selected
CC   from cytosine, thymine, and uracil and x = 6-16, and (ix) a combination
CC   of the primers. The method is useful for detecting genomic instability
CC   which are commonly associated with the various stages of neoplastic
CC   transformation and carcinogenesis. The method is rapid and simple.
XX
SQ   Sequence 10 BP; 0 A; 4 C; 5 G; 1 U; 0 other;

  Query Match             100.0%;  Score 8;  DB 20;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 5.3e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       1 CGCGCGCG 8


RESULT 36
AAX77472/c
ID   AAX77472 standard; DNA; 10 BP.
XX
AC   AAX77472;
XX
DT   05-AUG-1999  (first entry)
XX
DE   US5912147 primer 16.
XX
KW   Primer; quantitation; genetic instability; tumour cell; detection;
KW   neoplastic transformation; carcinogenesis; DNA/RNA hybrid; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   misc_RNA        10
FT                   /*tag=  a
FT                   /note= "uracil"
XX
PN   US5912147-A.
XX
PD   15-JUN-1999.
XX
PF   22-OCT-1996;   96US-0734973.
XX
PR   22-OCT-1996;   96US-0734973.
XX
PA   (HEAL-) HEALTH RES INC.
XX
PI   Anderson G,  Basik M,  Stoler D;
XX
DR   WPI; 1999-357197/30.
XX
PT   Quantitating genetic instability
XX
PS   Claim 4; Column 23-24; 27pp; English.
XX
CC   This invention describes a novel method for quantitating genetic
CC   instability independent of microsatellite alterations by treating a
CC   comparison pair comprising genomic DNA from tumour cells and genomic DNA
CC   from normal cells. The method involves the cells from the same individual
CC   with oligonucleotide primers selected from (i) a nucleotide sequence
CC   (CG)xRG, where R is a purine selected from adenine and guanine and x =
CC   3-7, (ii) a nucleotide sequence (CG)xRY, where R is as in (i) and Y is a
CC   pyrimidine selected from cytosine, thymine, and uracil and x = 3-7, (iii)
CC   a nucleotide sequence (CG)xRR, where R is as in (i) and x = 3-7, (iv) a
CC   nucleotide sequence (CG)xYY, where Y is a pyrimidine selected from
CC   cytosine, thymine, and uracil and x = 3-7, (v) a nucleotide sequence
CC   (CA)xRG, where R is a purine selected from adenine and guanine and x =
CC   6-16, (vi) a nucleotide sequence (CA)xRY, where R is a purine selected
CC   from adenine and guanine and Y is a pyrimidine  selected from cytosine,
CC   thymine, and uracil,  and x = 6-16, (vii) a nucleotide sequence (CA)xRR,
CC   where R is a purine selected from adenine and guanine and x = 6-16,
CC   (viii) a nucleotide sequence (CA)xYY, where Y is a pyrimidine selected
CC   from cytosine, thymine, and uracil and x = 6-16, and (ix) a combination
CC   of the primers. The method is useful for detecting genomic instability
CC   which are commonly associated with the various stages of neoplastic
CC   transformation and carcinogenesis. The method is rapid and simple.
XX
SQ   Sequence 10 BP; 0 A; 4 C; 5 G; 1 U; 0 other;

  Query Match             100.0%;  Score 8;  DB 20;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 5.3e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       8 CGCGCGCG 1


RESULT 37
AAX77473
ID   AAX77473 standard; DNA; 10 BP.
XX
AC   AAX77473;
XX
DT   05-AUG-1999  (first entry)
XX
DE   US5912147 primer 17.
XX
KW   Primer; quantitation; genetic instability; tumour cell; detection;
KW   neoplastic transformation; carcinogenesis; ss.
XX
OS   Synthetic.
XX
PN   US5912147-A.
XX
PD   15-JUN-1999.
XX
PF   22-OCT-1996;   96US-0734973.
XX
PR   22-OCT-1996;   96US-0734973.
XX
PA   (HEAL-) HEALTH RES INC.
XX
PI   Anderson G,  Basik M,  Stoler D;
XX
DR   WPI; 1999-357197/30.
XX
PT   Quantitating genetic instability
XX
PS   Claim 4; Column 23-24; 27pp; English.
XX
CC   This invention describes a novel method for quantitating genetic
CC   instability independent of microsatellite alterations by treating a
CC   comparison pair comprising genomic DNA from tumour cells and genomic DNA
CC   from normal cells. The method involves the cells from the same individual
CC   with oligonucleotide primers selected from (i) a nucleotide sequence
CC   (CG)xRG, where R is a purine selected from adenine and guanine and x =
CC   3-7, (ii) a nucleotide sequence (CG)xRY, where R is as in (i) and Y is a
CC   pyrimidine selected from cytosine, thymine, and uracil and x = 3-7, (iii)
CC   a nucleotide sequence (CG)xRR, where R is as in (i) and x = 3-7, (iv) a
CC   nucleotide sequence (CG)xYY, where Y is a pyrimidine selected from
CC   cytosine, thymine, and uracil and x = 3-7, (v) a nucleotide sequence
```

```
CC   (CA)xRG, where R is a purine selected from adenine and guanine and x =
CC   6-16, (vi) a nucleotide sequence (CA)xRY, where R is a purine selected
CC   from adenine and guanine and Y is a pyrimidine selected from cytosine,
CC   thymine, and uracil, and x = 6-16, (vii) a nucleotide sequence (CA)xRR,
CC   where R is a purine selected from adenine and guanine and x = 6-16,
CC   (viii) a nucleotide sequence (CA)xYY, where Y is a pyrimidine selected
CC   from cytosine, thymine, and uracil and x = 6-16, and (ix) a combination
CC   of the primers. The method is useful for detecting genomic instability
CC   which are commonly associated with the various stages of neoplastic
CC   transformation and carcinogenesis. The method is rapid and simple.
XX
SQ   Sequence 10 BP; 2 A; 4 C; 4 G; 0 U; 0 other;

  Query Match             100.0%;  Score 8;  DB 20;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 5.3e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       1 CGCGCGCG 8


RESULT 38
AAX77473/c
ID   AAX77473 standard; DNA; 10 BP.
XX
AC   AAX77473;
XX
DT   05-AUG-1999  (first entry)
XX
DE   US5912147 primer 17.
XX
KW   Primer; quantitation; genetic instability; tumour cell; detection;
KW   neoplastic transformation; carcinogenesis; ss.
XX
OS   Synthetic.
XX
PN   US5912147-A.
XX
PD   15-JUN-1999.
XX
PF   22-OCT-1996;   96US-0734973.
XX
PR   22-OCT-1996;   96US-0734973.
XX
PA   (HEAL-) HEALTH RES INC.
XX
PI   Anderson G,  Basik M,  Stoler D;
XX
DR   WPI; 1999-357197/30.
XX
PT   Quantitating genetic instability
XX
PS   Claim 4; Column 23-24; 27pp; English.
XX
CC   This invention describes a novel method for quantitating genetic
CC   instability independent of microsatellite alterations by treating a
CC   comparison pair comprising genomic DNA from tumour cells and genomic DNA
CC   from normal cells. The method involves the cells from the same individual
CC   with oligonucleotide primers selected from (i) a nucleotide sequence
CC   (CG)xRG, where R is a purine selected from adenine and guanine and x =
CC   3-7, (ii) a nucleotide sequence (CG)xRY, where R is as in (i) and Y is a
CC   pyrimidine selected from cytosine, thymine, and uracil and x = 3-7, (iii)
CC   a nucleotide sequence (CG)xRR, where R is as in (i) and x = 3-7, (iv) a
CC   nucleotide sequence (CG)xYY, where Y is a pyrimidine selected from
CC   cytosine, thymine, and uracil and x = 3-7, (v) a nucleotide sequence
CC   (CA)xRG, where R is a purine selected from adenine and guanine and x =
CC   6-16, (vi) a nucleotide sequence (CA)xRY, where R is a purine selected
CC   from adenine and guanine and Y is a pyrimidine selected from cytosine,
CC   thymine, and uracil, and x = 6-16, (vii) a nucleotide sequence (CA)xRR,
CC   where R is a purine selected from adenine and guanine and x = 6-16,
CC   (viii) a nucleotide sequence (CA)xYY, where Y is a pyrimidine selected
CC   from cytosine, thymine, and uracil and x = 6-16, and (ix) a combination
CC   of the primers. The method is useful for detecting genomic instability
CC   which are commonly associated with the various stages of neoplastic
CC   transformation and carcinogenesis. The method is rapid and simple.
XX
SQ   Sequence 10 BP; 2 A; 4 C; 4 G; 0 U; 0 other;

  Query Match             100.0%;  Score 8;  DB 20;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 5.3e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       8 CGCGCGCG 1


RESULT 39
AAX77474
ID   AAX77474 standard; DNA; 10 BP.
XX
AC   AAX77474;
XX
DT   05-AUG-1999  (first entry)
XX
DE   US5912147 primer 18.
XX
KW   Primer; quantitation; genetic instability; tumour cell; detection;
KW   neoplastic transformation; carcinogenesis; ss.
XX
OS   Synthetic.
XX
PN   US5912147-A.
XX
PD   15-JUN-1999.
XX
PF   22-OCT-1996;   96US-0734973.
XX
PR   22-OCT-1996;   96US-0734973.
XX
PA   (HEAL-) HEALTH RES INC.
XX
PI   Anderson G,  Basik M,  Stoler D;
XX
DR   WPI; 1999-357197/30.
XX
PT   Quantitating genetic instability
XX
PS   Claim 4; Column 23-24; 27pp; English.
XX
CC   This invention describes a novel method for quantitating genetic
CC   instability independent of microsatellite alterations by treating a
CC   comparison pair comprising genomic DNA from tumour cells and genomic DNA
CC   from normal cells. The method involves the cells from the same individual
CC   with oligonucleotide primers selected from (i) a nucleotide sequence
CC   (CG)xRG, where R is a purine selected from adenine and guanine and x =
CC   3-7, (ii) a nucleotide sequence (CG)xRY, where R is as in (i) and Y is a
CC   pyrimidine selected from cytosine, thymine, and uracil and x = 3-7, (iii)
CC   a nucleotide sequence (CG)xRR, where R is as in (i) and x = 3-7, (iv) a
CC   nucleotide sequence (CG)xYY, where Y is a pyrimidine selected from
CC   cytosine, thymine, and uracil and x = 3-7, (v) a nucleotide sequence
CC   (CA)xRG, where R is a purine selected from adenine and guanine and x =
CC   6-16, (vi) a nucleotide sequence (CA)xRY, where R is a purine selected
CC   from adenine and guanine and Y is a pyrimidine selected from cytosine,
CC   thymine, and uracil, and x = 6-16, (vii) a nucleotide sequence (CA)xRR,
CC   where R is a purine selected from adenine and guanine and x = 6-16,
CC   (viii) a nucleotide sequence (CA)xYY, where Y is a pyrimidine selected
CC   from cytosine, thymine, and uracil and x = 6-16, and (ix) a combination
CC   of the primers. The method is useful for detecting genomic instability
CC   which are commonly associated with the various stages of neoplastic
CC   transformation and carcinogenesis. The method is rapid and simple.
XX
SQ   Sequence 10 BP; 1 A; 4 C; 5 G; 0 U; 0 other;

  Query Match             100.0%;  Score 8;  DB 20;  Length 10;
```

```
  Best Local Similarity   100.0%;  Pred. No. 5.3e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db      1 CGCGCGCG 8


RESULT 40
AAX77474/c
ID   AAX77474 standard; DNA; 10 BP.
XX
AC   AAX77474;
XX
DT   05-AUG-1999  (first entry)
XX
DE   US5912147 primer 18.
XX
KW   Primer; quantitation; genetic instability; tumour cell; detection;
KW   neoplastic transformation; carcinogenesis; ss.
XX
OS   Synthetic.
XX
PN   US5912147-A.
XX
PD   15-JUN-1999.
XX
PF   22-OCT-1996;   96US-0734973.
XX
PR   22-OCT-1996;   96US-0734973.
XX
PA   (HEAL-) HEALTH RES INC.
XX
PI   Anderson G,  Basik M,  Stoler D;
XX
DR   WPI; 1999-357197/30.
XX
PT   Quantitating genetic instability
XX
PS   Claim 4; Column 23-24; 27pp; English.
XX
CC   This invention describes a novel method for quantitating genetic
CC   instability independent of microsatellite alterations by treating a
CC   comparison pair comprising genomic DNA from tumour cells and genomic DNA
CC   from normal cells. The method involves the cells from the same individual
CC   with oligonucleotide primers selected from (i) a nucleotide sequence
CC   (CG)xRG, where R is a purine selected from adenine and guanine and x =
CC   3-7, (ii) a nucleotide sequence (CG)xRY, where R is as in (i) and Y is a
CC   pyrimidine selected from cytosine, thymine, and uracil and x = 3-7, (iii)
CC   a nucleotide sequence (CG)xRR, where R is as in (i) and x = 3-7, (iv) a
CC   nucleotide sequence (CG)xYY, where Y is a pyrimidine selected from
CC   cytosine, thymine, and uracil and x = 3-7, (v) a nucleotide sequence
CC   (CA)xRG, where R is a purine selected from adenine and guanine and x =
CC   6-16, (vi) a nucleotide sequence (CA)xRY, where R is a purine selected
CC   from adenine and guanine and Y is a pyrimidine  selected from cytosine,
CC   thymine, and uracil,  and x = 6-16, (vii) a nucleotide sequence (CA)xRR,
CC   where R is a purine selected from adenine and guanine and x = 6-16,
CC   (viii) a nucleotide sequence (CA)xYY, where Y is a pyrimidine selected
CC   from cytosine, thymine, and uracil and x = 6-16, and (ix) a combination
CC   of the primers. The method is useful for detecting genomic instability
CC   which are commonly associated with the various stages of neoplastic
CC   transformation and carcinogenesis. The method is rapid and simple.
XX
SQ   Sequence 10 BP; 1 A; 4 C; 5 G; 0 U; 0 other;

  Query Match             100.0%;  Score 8;  DB 20;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 5.3e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGCGCG 8
          ||||||||
Db      8 CGCGCGCG 1


RESULT 41
AAA10396
ID   AAA10396 standard; DNA; 12 BP.
XX
AC   AAA10396;
XX
DT   03-JUL-2000  (first entry)
XX
DE   Kinetic strand displacement assay duplex competition oligo, SEQ ID NO:79.
XX
KW   Nucleic acid ligand binding assay; duplex formation; stability;
KW   detectable signal; kinetic strand displacement assay;
KW   competition oligonucleotide; ss.
XX
OS   Synthetic.
XX
PN   WO200015848-A1.
XX
PD   23-MAR-2000.
XX
PF   10-SEP-1999;   99WO-US20719.
XX
PR   11-SEP-1998;   98US-0151890.
XX
PA   (GENE-) GENELABS TECHNOLOGIES INC.
XX
PI   Schroth GP,  Bruice TW,  Suh YJ;
XX
DR   WPI; 2000-271478/23.
XX
PT   Determining binding affinity of a ligand to an oligonucleotide sequence
PT   in double stranded form, comprises measuring the effect of adding
PT   increasing amounts of a ligand on a signal generated by two indicator
PT   oligonucleotides of the duplex -
XX
PS   Example 12; Page 23; 78pp; English.
XX
CC   The invention relates to new methods of determining the binding affinity
CC   of a ligand to an oligonucleotide sequence, particularly to a duplex.
CC   The ligand is typically a metal ion, a small organic or inorganic
CC   molecule, a protein or a multi-protein complex. The methods comprise
CC   measuring the effect of adding increasing amounts of a ligand on a signal
CC   generated by two indicator oligonucleotides of the duplex. In the absence
CC   of ligand, conditions are such that the oligonucleotides exist primarily
CC   in single- stranded form; binding of ligand to double-stranded nucleic
CC   acids stabilises the duplexes, such that duplex formation is favoured.
CC   One of the indicator oligonucleotides contains a first group capable of
CC   producing a detectable signal, while the other indicator oligonucleotide
CC   contains a second group that on hybridisation of the two indicator
CC   molecules, will detectably alter the signal produced by the first group.
CC   The signal may be increased or decreased on hybridisation. For example,
CC   the pairs of signalling groups used could be a radioactive group and a
CC   scintillant (where an increase in signal intensity indicates that
CC   hybridisation has taken place) or a fluorophore and a fluorescence
CC   quencher (where a reduction in signal intensity indicates that
CC   hybridisation has occurred). Other methods of the invention comprise a
CC   strand diplacement assay, where the ability of an unlabelled
CC   displacement strand to displace one of the oligonucleotides in the duplex
CC   is determined in the absence and presence of ligand; and a competition
CC   assay, where an unlabelled single or double-stranded competitor
CC   oligonucleotide is added to the ligand-bound indicator duplex, and the
CC   effect on the signal produced from the indicator duplex determined. The
CC   methods are useful for determining the binding affinity of a ligand to
CC   an oligonucleotide sequence. They are particularly useful for
CC   determining relative binding affinities of various ligands to various
CC   oligonucleotide sequences, particularly double-stranded oligonucleotide
CC   sequences. The assays allow rapid and convenient determination of nucleic
CC   acid binding specificities. Sequences AAA10395-A10396 represent
CC   competition oligonucleotides used in a kinetic strand diplacement assay
CC   in an exemplification of the present invention.
XX
SQ   Sequence 12 BP; 0 A; 8 C; 4 G; 0 U; 0 other;
```

```
  Query Match             100.0%;  Score 8;  DB 21;  Length 12;
  Best Local Similarity   100.0%;  Pred. No. 5.1e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       3 CGCGCGCG 10


RESULT 42
AAA10396/c
ID   AAA10396 standard; DNA; 12 BP.
XX
AC   AAA10396;
XX
DT   03-JUL-2000  (first entry)
XX
DE   Kinetic strand displacement assay duplex competition oligo, SEQ ID NO:79.
XX
KW   Nucleic acid ligand binding assay; duplex formation; stability;
KW   detectable signal; kinetic strand displacement assay;
KW   competition oligonucleotide; ss.
XX
OS   Synthetic.
XX
PN   WO200015848-A1.
XX
PD   23-MAR-2000.
XX
PF   10-SEP-1999;   99WO-US20719.
XX
PR   11-SEP-1998;   98US-0151890.
XX
PA   (GENE-) GENELABS TECHNOLOGIES INC.
XX
PI   Schroth GP,  Bruice TW,  Suh YJ;
XX
DR   WPI; 2000-271478/23.
XX
PT   Determining binding affinity of a ligand to an oligonucleotide sequence
PT   in double stranded form, comprises measuring the effect of adding
PT   increasing amounts of a ligand on a signal generated by two indicator
PT   oligonucleotides of the duplex -
XX
PS   Example 12; Page 23; 78pp; English.
XX
CC   The invention relates to new methods of determining the binding affinity
CC   of a ligand to an oligonucleotide sequence, particularly to a duplex.
CC   The ligand is typically a metal ion, a small organic or inorganic
CC   molecule, a protein or a multi-protein complex. The methods comprise
CC   measuring the effect of adding increasing amounts of a ligand on a signal
CC   generated by two indicator oligonucleotides of the duplex. In the absence
CC   of ligand, conditions are such that the oligonucleotides exist primarily
CC   in single- stranded form; binding of ligand to double-stranded nucleic
CC   acids stabilises the duplexes, such that duplex formation is favoured.
CC   One of the indicator oligonucleotides contains a first group capable of
CC   producing a detectable signal, while the other indicator oligonucleotide
CC   contains a second group that on hybridisation of the two indicator
CC   molecules, will detectably alter the signal produced by the first group.
CC   The signal may be increased or decreased on hybridisation. For example,
CC   the pairs of signalling groups used could be a radioactive group and a
CC   scintillant (where an increase in signal intensity indicates that
CC   hybridisation has taken place) or a fluorophore and a fluorescence
CC   quencher (where a reduction in signal intensity indicates that
CC   hybridisation has occurred). Other methods of the invention comprise a
CC   strand diplacement assay, where the ability of an unlabelled
CC   displacement strand to displace one of the oligonucleotides in the duplex
CC   is determined in the absence and presence of ligand; and a competition
CC   assay, where an unlabelled single or double-stranded competitor
CC   oligonucleotide is added to the ligand-bound indicator duplex, and the
CC   effect on the signal produced from the indicator duplex determined. The
CC   methods are useful for determining the binding affinity of a ligand to
CC   an oligonucleotide sequence. They are particularly useful for
CC   determining relative binding affinities of various ligands to various
CC   oligonucleotide sequences, particularly double-stranded oligonucleotide
CC   sequences. The assays allow rapid and convenient determination of nucleic
CC   acid binding specificities. Sequences AAA10395-A10396 represent
CC   competition oligonucleotides used in a kinetic strand diplacement assay
CC   in an exemplification of the present invention.
XX
SQ   Sequence 12 BP; 0 A; 8 C; 4 G; 0 U; 0 other;

  Query Match             100.0%;  Score 8;  DB 21;  Length 12;
  Best Local Similarity   100.0%;  Pred. No. 5.1e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db      10 CGCGCGCG 3


RESULT 43
AAS13767
ID   AAS13767 standard; DNA; 12 BP.
XX
AC   AAS13767;
XX
DT   08-MAY-2002  (first entry)
XX
DE   Simple sequence repeat, SSR, #39.
XX
KW   Simple sequence repeat; plant; ds; SSR; ryegrass; fescue; tandem repeat;
KW   cereal profiling; grass profiling; seed batch purity testing.
XX
OS   Festuca aruninacea.
XX
PN   NZ509193-A.
XX
PD   25-MAY-2001.
XX
PF   03-JAN-2001; 2001NZ-0509193.
XX
PR   24-DEC-1999;   99AU-0004906.
PR   04-MAY-2000; 2000AU-0007310.
XX
PA   (SAUS-) STATE SOUTH AUSTRALIA SOUTH AUSTRALIAN R.
PA   (UYSC-) UNIV SOUTHERN CROSS.
PA   (VICT-) STATE VICTORIA DEPT NATURAL RES & ENVIRO.
PA   (UYAD-) UNIV ADELAIDE.
PA   (ITMA-) INT MAIZE & WHEAT IMPROVEMENT CENT.
XX
PI   Forster JW,  Jones ES;
XX
DR   WPI; 2001-512563/56.
XX
PT   New simple sequence repeats having 2 or more tandemly repeated
PT   nucleotide core elements isolated from ryegrass and fescue, useful for
PT   selecting of genes in grass or cereal breeding or profiling grass or
PT   cereal species varieties  -
XX
PS   Example 1; Fig 6; 72pp; English.
XX
CC   The invention relates to a substantially purified or isolated nucleic
CC   acid (I) from ryegrass or fescue species including a simple sequence
CC   repeat (SSR), having 2 or more tandemly repeated nucleotide core elements
CC   2-6 nucleotides in length. Also included are a nucleic acid primer
CC   suitable for amplifying an SSR, identifying (M1) an SSR by preparing a
CC   library of ryegrass or fescue genomic DNA enriched for SSRs and
CC   identifying clones in the library containing SSRs, a library of ryegrass
CC   or fescue genomic DNA enriched for SSRs prepared by the M1, selecting for
CC   a gene in grass or cereal breeding by identifying an SSR that is closely
CC   associated with the gene such that the SSR and the gene are
CC   preferentially co-inherited, and selecting for the SSR in the
CC   breeding, a method for DNA profiling grass or cereal species varieties by
CC   assessing variation between SSR varieties and testing the purity of grass
```

```
CC   or cereal seed batches by assessing variation within seed batch of an
CC   SSR. The SSRs may be used in the selection of genes in grass or cereal
CC   breeding, for profiling grass or cereal species varieties, for testing
CC   the purity of grass or cereal seed batches, and for DNA profiling to
CC   establish the distinct identity, uniformity and/or stability of a
CC   cultivar. The present sequence is a ryegrass or fescue SSR.
XX
SQ   Sequence 12 BP; 0 A; 5 C; 6 G; 1 T; 0 other;

  Query Match             100.0%;  Score 8;  DB 23;  Length 12;
  Best Local Similarity   100.0%;  Pred. No. 5.1e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db       4 CGCGCGCG 11


RESULT 44
AAS13767/c
ID   AAS13767 standard; DNA; 12 BP.
XX
AC   AAS13767;
XX
DT   08-MAY-2002  (first entry)
XX
DE   Simple sequence repeat, SSR, #39.
XX
KW   Simple sequence repeat; plant; ds; SSR; ryegrass; fescue; tandem repeat;
KW   cereal profiling; grass profiling; seed batch purity testing.
XX
OS   Festuca aruninacea.
XX
PN   NZ509193-A.
XX
PD   25-MAY-2001.
XX
PF   03-JAN-2001; 2001NZ-0509193.
XX
PR   24-DEC-1999;   99AU-0004906.
PR   04-MAY-2000; 2000AU-0007310.
XX
PA   (SAUS-) STATE SOUTH AUSTRALIA SOUTH AUSTRALIAN R.
PA   (UYSC-) UNIV SOUTHERN CROSS.
PA   (VICT-) STATE VICTORIA DEPT NATURAL RES & ENVIRO.
PA   (UYAD-) UNIV ADELAIDE.
PA   (ITMA-) INT MAIZE & WHEAT IMPROVEMENT CENT.
XX
PI   Forster JW,  Jones ES;
XX
DR   WPI; 2001-512563/56.
XX
PT   New simple sequence repeats having 2 or more tandemly repeated
PT   nucleotide core elements isolated from ryegrass and fescue, useful for
PT   selecting of genes in grass or cereal breeding or profiling grass or
PT   cereal species varieties  -
XX
PS   Example 1; Fig 6; 72pp; English.
XX
CC   The invention relates to a substantially purified or isolated nucleic
CC   acid (I) from ryegrass or fescue species including a simple sequence
CC   repeat (SSR), having 2 or more tandemly repeated nucleotide core elements
CC   2-6 nucleotides in length. Also included are a nucleic acid primer
CC   suitable for amplifying an SSR, identifying (M1) an SSR by preparing a
CC   library of ryegrass or fescue genomic DNA enriched for SSRs and
CC   identifying clones in the library containing SSRs, a library of ryegrass
CC   or fescue genomic DNA enriched for SSRs prepared by the M1, selecting for
CC   a gene in grass or cereal breeding by identifying an SSR that is closely
CC   associated with the gene such that the SSR and the gene are
CC   preferentially co-inherited, and selecting for the SSR in the
CC   breeding, a method for DNA profiling grass or cereal species varieties by
CC   assessing variation between SSR varieties and testing the purity of grass
CC   or cereal seed batches by assessing variation within seed batch of an
CC   SSR. The SSRs may be used in the selection of genes in grass or cereal
CC   breeding, for profiling grass or cereal seed batches, and for DNA profiling to
CC   the purity of grass or cereal species varieties, for testing
CC   establish the distinct identity, uniformity and/or stability of a
CC   cultivar. The present sequence is a ryegrass or fescue SSR.
XX
SQ   Sequence 12 BP; 0 A; 5 C; 6 G; 1 T; 0 other;

  Query Match             100.0%;  Score 8;  DB 23;  Length 12;
  Best Local Similarity   100.0%;  Pred. No. 5.1e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
           ||||||||
Db      11 CGCGCGCG 4


RESULT 45
ABH86501
ID   ABH86501 standard; DNA; 12 BP.
XX
AC   ABH86501;
XX
DT   22-FEB-2002  (first entry)
XX
DE   Oligonucleotide primer SEQ ID NO 286494 for detecting SNP TSC0012735.
XX
KW   SNP; single nucleotide polymorphism; human; diagnosis; PNA; cancer; CNS;
KW   peptide nucleic acid; cytosine methylation; cardiovascular; primer; ss;
KW   central nervous system; gastrointestinal; respiratory; immune; metabolic.
XX
OS   Homo sapiens.
XX
PN   WO200177384-A2.
XX
PD   18-OCT-2001.
XX
PF   06-APR-2001; 2001WO-IB00713.
XX
PR   07-APR-2000; 2000DE-1019173.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2001-657177/75.
XX
PT   Set of oligonucleotides, useful for diagnosis and cell typing, is
PT   designed to detect single nucleotide polymorphisms and cytosine
PT   methylation status  -
XX
PS   Claim 1; SEQ ID 286494; 29pp + Sequence Listing; German.
XX
CC   This invention describes novel oligonucleotide primers or peptide nucleic
CC   acid (PNA) oligomers for detecting single nucleotide polymorphisms (SNP)
CC   and cytosine methylation status in chemically pretreated genomic DNA. The
CC   oligonucleotides are used for diagnosis and/or prognosis of cancer and a
CC   range of diseases including immune system, gastrointestinal, respiratory,
CC   central nervous system, cardiovascular and metabolic disorders. The
CC   oligomers are also used for detecting cell type differentiation.
CC   ABC00010-ABC99989, ABF00010-ABF99989, ABH00010-ABH99989 and
CC   ABI00010-ABI82073 represent the oligomers described in the invention.
CC   NOTE: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 12 BP; 0 A; 4 C; 6 G; 2 T; 0 other;

  Query Match             100.0%;  Score 8;  DB 23;  Length 12;
  Best Local Similarity   100.0%;  Pred. No. 5.1e+04;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGCGCG 8
```

```
          ||||||||
Db        1 CGCGCGCG 8

Search completed: March 14, 2003, 03:41:37
Job time : 198 secs
```